

簡易的大気大循環モデルを用いた
ハイブリッドデータ同化に関する研究

2017年1月

寺内俊平

簡易的大気大循環モデルを用いた ハイブリッドデータ同化に関する研究

筑波大学大学院
生命環境科学研究科
地球科学専攻
修士（理学）学位論文

寺内俊平

A Study on Hybrid Data Assimilation with a Simplified Primitive-Equation Global Model

Shumpei TERAUCHI

Abstract

Four Dimension VARIational assimilation (4DVAR) and Ensemble Kalman Filter (EnKF) have been mainly studied in the field of meteorology. Recently, a hybrid of these methods is receiving increasing interest.

In this study, Hybrid-3DVAR was focused on as one of the hybrid data assimilation methods. The characteristics with a method of regenerating analysis ensemble of the method was examined. No-cost Ensemble Kalman Smoother (No-cost EnKS) is one of the smoother methods using Local Ensemble Transform Kalman Filter (LETKF) analysis. The purpose of this study is to investigate how well improved the performance of the data assimilation by using th Hybrid-3DVAR with No-cost EnKS.

According to the results of comparison between pure-LETKF and No-cost EnKS, a performance of No-cost EnKS with only one iteration showed higher than that of pure-LETKF. However, the performance of No-cost EnKS experiment with two or more iterations showed lower than that of one iteretion.

In conclusion, the Hybrid-3DVAR was developed and compared to the pure-3DVAR in this study. As a result, the performance of Hybrid-3DVAR outperformed that of pure-3DVAR. The Hybrid-3DVAR with No-cost EnKS, however, has slight impact on reducing analysis error.

Key words : Data assimilation, Hybrid data assimilation, No-cost ensemble Kalman smoother, Running in place scheme

目次

Abstract	i
図目次	iii
1 はじめに	1
2 目的	3
3 同化手法	4
3.1 線形最小分散推定と最尤推定	4
3.1.1 線形最小分散推定	4
3.1.2 最尤推定	6
3.2 カルマンフィルタ	7
3.3 アンサンブルカルマンフィルタ	10
3.4 変分法	13
3.4.1 3次元変分法	14
3.4.2 前処理	15
3.5 ハイブリッドデータ同化	15
3.6 安定化	15
3.6.1 共分散膨張	16
3.6.2 局所化	16
3.7 No-cost Ensemble Kalman Smoother と RIP スキーム	16
4 実験設定	18
4.1 SPEEDY	18
4.1.1 実験1 LETKF, No-cost EnKS および RIP スキームの比較実験	18
4.1.2 実験2 3DVAR および Hybrid-3DVAR の比較実験	19
5 結果	20
5.1 実験1の結果	20
5.2 実験2の結果	20
6 考察	22
6.1 実験1の考察	22
6.2 実験2の考察	23
7 結論	24
謝辞	25
参考文献	26

目 次

1	アンサンブルカルマンフィルタ (EnKF) の概念図.	28
2	3次元変分法 (3DVAR) の概念図. 灰色の矢印が同化を行う前の予報値のトラジェクトリを表している.	29
3	4次元変分法 (4DVAR) の概念図. 灰色の矢印が同化を行う前の予報値のトラジェクトリを表している.	30
4	ハイブリッドデータ同化の概念図. \mathbf{B}_{nmc} は NMC 法による統計的な背景誤差共分散, \mathbf{B}_{ens} はアンサンブルから求めた流れ依存性を持つ背景誤差共分散を表す.	31
5	本実験で用いた観測分布. 黒色の点が観測地点を表している. 観測密度は 9.03 % となっている.	32
6	鉛直第 4 層における気温の解析二乗平均平方根誤差 (RMSE). 縦軸が RMSE の値で, 横軸が時系列を表している. それぞれ, 緑色の円が通常の LETKF, 赤色の円が 1 回の繰り返しの No-cost EnKS, 青色の円が 2 回の繰り返しの No-cost EnKS, 水色の円が 3 回の繰り返しの No-cost EnKS, 橙色の四角形が RIP スキームによる結果を示している.	33
7	鉛直第 4 層における東西風の解析二乗平均平方根誤差 (RMSE). それ以外は図 6 に同じ.	34
8	鉛直第 2 層における比湿の解析二乗平均平方根誤差 (RMSE). それ以外は図 6 に同じ.	35
9	地表面気圧の解析二乗平均平方根誤差 (RMSE). それ以外は図 6 に同じ.	36
10	全球での 6 時間積算降水量の真値に対する相関係数. 凡例等は図 6 に同じ.	37
11	1982 年 1 月 15 日 12 時における鉛直第 4 層における LETKF の解析の気温の RMSE. 陰影が RMSE の値を表しており, 暖色系ほど値が大きいことを示す.	38
12	1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 1 回) の解析の気温の RMSE. 陰影の設定は図 11 と同じ.	39
13	1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 2 回) の解析の気温の RMSE. 陰影の設定は図 11 と同じ.	40
14	1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 3 回) の解析の気温の RMSE. 陰影の設定は図 11 と同じ.	41
15	1982 年 1 月 15 日 12 時における鉛直第 4 層における LETKF の解析の気温の Spread. 陰影が spread の値を表しており, 暖色系ほど値が大きいことを示す.	42
16	1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 1 回) の解析の気温の Spread. 陰影の設定は図 15 に同じ.	43
17	1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 2 回) の解析の気温の Spread. 陰影の設定は図 15 に同じ.	44
18	1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 3 回) の解析の気温の Spread. 陰影の設定は図 15 に同じ.	45
19	1982 年 1 月 15 日 12 時における鉛直第 4 層における LETKF の解析 spread と No-cost EnKS (繰り返し 1 回) の解析 spread の差分. 陰影は暖色系になるほど LETKF のほうが spread が大きいことを示し, 寒色系になるほど No-cost EnKS のほうが spread が大きいことを示す.	46

20	1982年1月15日12時における鉛直第4層における LETKF の解析 spread と No-cost EnKS (繰返し2回) の解析 spread の差分. 陰影の設定は図19に同じ.	47
21	1982年1月15日12時における鉛直第4層における LETKF の解析 spread と No-cost EnKS (繰返し3回) の解析 spread の差分. 陰影の設定は図19に同じ.	48
22	NMC法で作成した, 第4層における背景誤差の東西風の標準偏差.	49
23	NMC法で作成した, 第4層における背景誤差の気温の標準偏差.	50
24	1982年1月15日12時における解析アンサンブルで作成した, 第4層における背景誤差の東西風の標準偏差.	51
25	1982年1月15日12時における解析アンサンブルで作成した, 第4層における背景誤差の気温の標準偏差.	52
26	鉛直第4層における気温の解析 RMSE. 縦軸が RMSE の値で, 横軸が時系列を表している. それぞれ, 緑色の円が通常の LETKF, 赤色の円が通常の LETKF による解析アンサンブルによる背景誤差のみを用いた 3DVAR, 青色の円が Hybrid-3DVAR, 水色の円が通常の NMC 法による背景誤差のみを用いた 3DVAR, 橙色の四角形が No-cost EnKS による解析アンサンブルを用いた Hybrid-3DVAR, 紫色の四角形が LETKF のメンバー数を 80 に増やしたときの Hybrid-3DVAR を示す.	53
27	鉛直第4層における東西風の解析 RMSE. 縦軸が RMSE の値で, 横軸が時系列を表している. 凡例は図26に同じ.	54

1 はじめに

近年の天気予報において、数値予報は重要な地位を占めている。この数値予報の精度を向上させる方法は大きく分けて2つあり、1つは数値予報モデル自体の精緻化を図ることであり、もう1つはデータ同化手法を高度化させることである。

上述のデータ同化とは、簡単に言うと、観測値や実験データを線形最小分散推定や最尤推定といった推定法に基づいた上で、最適化理論を応用してうまく取り込む、すなわち、同化することによってモデル結果を修正する技法のことである(淡路ほか 2009)。現在、気象分野のデータ同化の手法で主に用いられている手法としては、前述した線形最小分散推定を基に構築されているアンサンブルカルマンフィルタ(Ensemble Kalman Filter: EnKF)と、最尤推定法の考え方を利用している4次元変分法(Four Dimensional Variational Assimilation: 4DVAR)の2つが挙げられる。この2つの手法は誤差の情報の時間発展を考慮しているという点では共通しているが、EnKFは前方にしか観測情報を伝搬しないのに対し、4DVARはバックワード演算を行って、観測情報を後方にも伝搬していくという点で異なる。この点だけを考慮すると、4DVARの方が優れていることが推測される。しかしながら、EnKFはフォワードモデルのみを考慮すれば良いのに対し、4DVARはバックワード演算のために接線形モデルと呼ばれるモデル演算子を線形化したもの、およびその転置形であるアジョイントモデルをコーディングしなければならない。この点を加味すると、マンパワーを用意できないような気象予報機関にとって4DVARの実装は困難であると考えられる。その上、接線形モデルおよびアジョイントモデルは数値モデルに依存しているため、数値モデルの更新の際にこれらのモデルも同時に更新する必要があり、マンパワーがここにも必要になってしまうことも考えられる。そのため、実装という面で比較した場合は4DVARよりもEnKFの方が有利である。

また、今までに4DVARとEnKFの精度の比較は数多く行われており、理想的な状況下、すなわちモデルの線形性および完全性が保証されていて、4DVARについては同化ウィンドウの長さを無限長とし、EnKFについてはアンサンブルサイズを無限大としたとき、最終的な精度は双方とも等しくなることが知られている(例えば、Bouttier and Courtier 1999)。実際のモデルの同化システムの精度に関しては、様々な視点から議論が行われてはいるものの(例えば、Kalnay et al. 2007a)、どちらが有利であるかということに関しては、個々のモデルの不確実性などに依存するので、未だに結論は出ていない状況である。

しかしながら、近年では4DVARとEnKFの精度の比較研究よりも、4DVARとEnKFを融合させた手法の研究が盛んになってきており、その中の一つにハイブリッドデータ同化と呼ばれるものがある。この手法はHamill and Snyder (2000)によって初めて考案された手法であり、その特徴は、従来の変分法の短所であった統計的な、すなわち流れ依存性を考慮できない背景誤差を、アンサンブル予報による背景誤差と加重平均をとることによって流れ依存性をもつ背景誤差を変分法に取り込むといった点にある。変分法の同化システムにおいて、背景誤差の情報は同化後の精度に大きく影響するため、アンサンブル予報によって作成する背景誤差の精度自身もハイブリッドデータ同化の精度に影響すると考えられる。

そこで、本研究ではEnKFの解析アンサンブルを作り直す手法であるNo-cost Ensemble Kalman Smoother(EnKS: Yang et al. 2009)およびrunning in place(RIP)スキーム(Kalnay and Yang 2010; Yang et al. 2012)という手法をハイブリッドデータ同化と併用させることによる影響の調査を試みた。

4DVARとEnKFおよびハイブリッドデータ同化に関しては、前述の通り数多くの比較研

究が行われている (例えば, Lorenc 2003; Wang et al. 2008) ものの, ハイブリッドデータ同化と EnKS および RIP スキームを組み合わせた先行研究はない. そのため, ハイブリッドデータ同化における作り直された解析アンサンブルによる背景誤差がどの程度影響するのかという点について調査することは, ハイブリッドデータ同化手法の研究の発展に寄与し得るはずである.

本論文では, 第 2 章で本研究の目的を改めて整理した後, 第 3 章では本研究で構築したデータ同化手法 EnKF, 3DVAR, およびハイブリッドデータ同化手法の導出・説明を行う. 第 4 章では, 簡易的な大気大循環モデルである SPEEDY の簡単な説明およびその設定を述べた後に, 第 5 章でその実験結果をまとめ, 第 6 章に考察を記した.

2 目的

本研究では、簡易的な大気大循環モデル SPEEDY を用いて、ハイブリッドデータ同化手法にアンサンブルを作り直す手法である No-cost EnKS および RIP スキームを実装したときの影響について調査することを目的とする。

3 同化手法

以下、淡路ほか (2009), 近藤 (2009), 露木・川畑 (2008), 堂山 (2014), 三好 (2005, 2006) を参考にして, 同化手法の説明を行う。

3.1 線形最小分散推定と最尤推定

3.1.1 線形最小分散推定

線形最小分散推定とは, ある推定値を状態変数の線形結合で表現したとき, その誤差分散を最小にするような重みを求める推定法のことである。以下, 簡単な数式を用いて説明する。

ここでは, 簡単のため, 状態変数は異なる2つの推定値 (変数) x_1, x_2 しかない場合を考える。ここで, 変数の真値を x^t , 2つの推定値の誤差を $\varepsilon_1, \varepsilon_2$ とすると, 推定値 x_1, x_2 は以下のように表すことができる。

$$x_1 = x^t + \varepsilon_1 \quad (3.1.1)$$

$$x_2 = x^t + \varepsilon_2 \quad (3.1.2)$$

また, この推定法では, 次の2つの仮定をおくことによって最適な推定値を求める。

1. x_1, x_2 はバイアスを持たない値, すなわち不偏推定値である。期待値を $\langle \cdot \rangle$ で表すとしたとき, x_1, x_2 および $\varepsilon_1, \varepsilon_2$ は次の関係を満たす。

$$\langle x_1 \rangle = \langle x_2 \rangle = x^t \quad (3.1.3)$$

$$\langle \varepsilon_1 \rangle = \langle \varepsilon_2 \rangle = 0 \quad (3.1.4)$$

2. x_1 と x_2 の誤差は無相関である。このことを式で表現すると,

$$\langle (x_1 - x^t)(x_2 - x^t) \rangle = 0 \quad (3.1.5)$$

となる。(3.1.1) 式および (3.1.2) 式の間係を用いると (3.1.5) 式は,

$$\langle \varepsilon_1 \varepsilon_2 \rangle = 0 \quad (3.1.6)$$

とも, 表現できる。

以上の仮定の下で, 最適推定値 x_a を考える。最適推定値 x_a を, x_1, x_2 の線形結合で表すと,

$$x_a = \alpha_1 x_1 + \alpha_2 x_2 \quad (3.1.7)$$

となる。この x_a の誤差分散 σ_a^2 を最小にする係数 (重み) α_1, α_2 を求めることが, 前述した線形最小分散推定と呼ばれる推定法である。

x_a も不偏推定量だと仮定すると,

$$\langle x_a \rangle = x^t \quad (3.1.8)$$

が成り立つ必要がある。ここで, (3.1.8) 式に (3.1.7) 式を代入すると,

$$\begin{aligned} \langle \alpha_1 x_1 + \alpha_2 x_2 \rangle &= x^t \\ \alpha_1 \langle x_1 \rangle + \alpha_2 \langle x_2 \rangle &= x^t \end{aligned} \quad (3.1.9)$$

となる。(3.1.3) 式の関係より,

$$\begin{aligned}\alpha_1 x^t + \alpha_2 x^t &= x^t \\ \{(\alpha_1 + \alpha_2) - 1\} x^t &= 0 \\ \alpha_1 + \alpha_2 &= 1\end{aligned}\tag{3.1.10}$$

を得る。(3.1.10) 式の関係より, (3.1.7) 式は次のようにも表現できる.

$$x_a = \alpha_1 x_1 + (1 - \alpha_1) x_2\tag{3.1.11}$$

以上の関係から, x_a の誤差分散 σ_a^2 は次のように表すことができる.

$$\begin{aligned}\sigma_a^2 &= \langle (x_a - x^t)^2 \rangle \\ &= \alpha_1^2 \langle \varepsilon_1^2 \rangle + (1 - \alpha_1)^2 \langle \varepsilon_2^2 \rangle \\ &= \alpha_1^2 \sigma_1^2 + (1 - \alpha_1)^2 \sigma_2^2\end{aligned}\tag{3.1.12}$$

ここで, x_1, x_2 の誤差分散をそれぞれ σ_1^2, σ_2^2 とした.

$$\langle (x_1 - x^t)^2 \rangle = \langle \varepsilon_1^2 \rangle = \sigma_1^2\tag{3.1.13}$$

$$\langle (x_2 - x^t)^2 \rangle = \langle \varepsilon_2^2 \rangle = \sigma_2^2\tag{3.1.14}$$

σ_a^2 が最小であるためには, $\frac{\partial \sigma_a^2}{\partial \alpha_1} = 0$ を満たすような α_1 を求めれば良い.
(3.1.12) 式を α_1 で偏微分すると,

$$2\alpha_1 \sigma_1^2 - 2(1 - \alpha_1) \sigma_2^2 = 0$$

$$2\alpha_1 \sigma_1^2 - 2\sigma_2^2 + 2\alpha_1 \sigma_2^2 = 0$$

となる. この上式を α_1 について解くと,

$$\alpha_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}\tag{3.1.15}$$

を得る. また, (3.1.10) 式を用いると,

$$\alpha_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}\tag{3.1.16}$$

を得る.

得られた (3.1.15) 式, (3.1.16) 式を用いると, x_a および σ_a^2 は次のように表現できる.

$$\begin{aligned}x_a &= \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} x_2 \\ &= x_1 + \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2} (x_2 - x_1)\end{aligned}\tag{3.1.17}$$

$$\sigma_a^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}\tag{3.1.18}$$

この (3.1.17) 式が求める最適推定値 x_a である.

3.1.2 最尤推定

3.1.1 節と同様に、2つの推定値 x_1, x_2 から最適推定値 x_n を求めることを考える。ここでは、 x_1, x_2 は以下の仮定に従う確率変数とおく。

- 1) x_1, x_2 は正規分布に従う。
- 2) x_1, x_2 は不偏推定値である。すなわち、期待値は真値に等しい。
- 3) x_1, x_2 は独立である。

仮定 1), 2) から x_1, x_2 の確率密度分布は次のように書ける。

$$p(x_1|x^t) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left[-\frac{(x_1 - x^t)^2}{2\sigma_1^2} \right] \quad (3.1.19)$$

$$p(x_2|x^t) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left[-\frac{(x_2 - x^t)^2}{2\sigma_2^2} \right] \quad (3.1.20)$$

このとき、 x_1, x_2 が同時に実現する確率密度分布 $p(x_1, x_2|x^t)$ は仮定 3) より、(3.1.19) 式と (3.1.20) 式の積で表現できる。これを $L(x^t|x_1, x_2)$ と表記する。

$$\begin{aligned} p(x_1, x_2|x^t) &= p(x_1|x^t)p(x_2|x^t) \\ &= L(x^t|x_1, x_2) \end{aligned} \quad (3.1.21)$$

ここで、未知である真値を変数と見て x と表記し、その最適推定値 x_n を求めることを考える。このとき $L(x|x_1, x_2)$ は以下のように書ける。

$$L(x|x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2} \exp \left[-\frac{(x_1 - x^t)^2}{2\sigma_1^2} - \frac{(x_2 - x^t)^2}{2\sigma_2^2} \right] \quad (3.1.22)$$

(3.1.22) 式は、ある真値 x を仮定したとき、 x_1, x_2 というデータが同時に観測される確率を表す。言い換えると、 x_1, x_2 の確率分布の前提となる真値 x の尤もらしさを表している。そのため、この L は一般に尤度関数と呼ばれる。

次に、 L を最大にする x を求める最適化問題について考える。 L を最大にするためには、指数関数の引数が最大になればよいので、引数にマイナスを掛けた次式を最小にすればよい。

$$J(x) = \frac{(x_1 - x^t)^2}{2\sigma_1^2} + \frac{(x_2 - x^t)^2}{2\sigma_2^2} \quad (3.1.23)$$

以上のような推定法を最尤推定と呼ぶ。また、(3.1.23) 式の J は評価関数、その変数 x は制御変数を表す。

(3.1.23) 式を見てわかるように、評価関数 J は推定値 x_1, x_2 と推定値 x の差の二乗をそれぞれの誤差分散で規格化したものになっている。

3.2 カルマンフィルタ

カルマンフィルタ (KF) とは, Kalman (1960) によって提唱された最適制御理論である. この理論は誤差の時間発展が線形性を持ち, 誤差の確率分布が正規分布であるという仮定の下で最適解を求めるというものである.

この節では, 前節の線形最小分散推定の考え方を元にして, カルマンフィルタの式の導出を行う.

まず, 時空間的に離散化された線形の力学モデルを考える. ただし, モデルは完全であり, システムノイズは発生しないものとする (強拘束条件).

$$\mathbf{x}_t^f = \mathbf{M}\mathbf{x}_{t-1}^a \quad (3.2.1)$$

ここで, \mathbf{x} は状態ベクトルを表し, \mathbf{M} は状態遷移行列, すなわち線形化された予報モデルを表す. また, 上付きの添字 f , a はそれぞれ予報 (forecast), 解析 (analysis) を表し, 下付きの添字は時刻を表している.

次に, 状態ベクトルの誤差について考える. (3.1.1) 節で導入した推定値の誤差の議論を一般化すると, ある時間ステップ t における予報値の状態ベクトル \mathbf{x}_t^f と解析値の状態ベクトル \mathbf{x}_t^a は以下のように表せる.

$$\mathbf{x}_t^f = \mathbf{x}_t^t + \mathbf{p}_t^f \quad (3.2.2)$$

$$\mathbf{x}_t^a = \mathbf{x}_t^t + \mathbf{p}_t^a \quad (3.2.3)$$

ここで, 上付きの添字 t は真値を意味し, ベクトル \mathbf{p} は誤差ベクトルを意味する. (3.2.1), (3.2.2) および (3.2.3) 式を用いると, 予報誤差共分散行列 \mathbf{P}^f は次のように与えられる.

$$\begin{aligned} \mathbf{P}_t^f &= \langle \mathbf{p}_t^f (\mathbf{p}_t^f)^T \rangle = \langle (\mathbf{x}_t^f - \mathbf{x}_t^t) (\mathbf{x}_t^f - \mathbf{x}_t^t)^T \rangle \\ &= \langle (\mathbf{M}\mathbf{x}_{t-1}^a - \mathbf{M}\mathbf{x}_{t-1}^t) (\mathbf{M}\mathbf{x}_{t-1}^a - \mathbf{M}\mathbf{x}_{t-1}^t)^T \rangle \\ &= \langle (\mathbf{M}(\mathbf{x}_{t-1}^a - \mathbf{x}_{t-1}^t)) (\mathbf{M}(\mathbf{x}_{t-1}^a - \mathbf{x}_{t-1}^t))^T \rangle \\ &= \langle (\mathbf{M}\mathbf{p}_{t-1}^a) (\mathbf{M}\mathbf{p}_{t-1}^a)^T \rangle \\ &= \langle \mathbf{M}\mathbf{p}_{t-1}^a (\mathbf{p}_{t-1}^a)^T \mathbf{M}^T \rangle \\ &= \mathbf{M}\mathbf{P}_{t-1}^a \mathbf{M}^T \end{aligned} \quad (3.2.4)$$

ここで, 強拘束条件により真値の時間発展は真値を与えるものとし, 解析誤差共分散行列 \mathbf{P}^a は予報誤差共分散行列と同様に誤差ベクトルのクロスタームで与えられるものとした. この (3.2.4) 式は, 同化が全く行われない場合における予報誤差の時間発展を表しており, リヤプノフ方程式という名前がついている.

次に観測を同化させる場合の, 予報誤差の時間発展を考える. 時間 t における観測値 (観測ベクトル) \mathbf{y}_t^o は, 以下のように表すことができる.

$$\mathbf{y}_t^o = \mathbf{H}_t \mathbf{x}_t^t + \mathbf{r}_t \quad (3.2.5)$$

ここで, \mathbf{H} は観測行列, \mathbf{r} は観測誤差ベクトルである. \mathbf{H} は, 状態空間から観測空間への射影を行う行列である. また, 上付きの添字 o は観測を表す.

同化後の解析値（最適推定値）は，前節の線形最小分散推定に基づき，予測値 \mathbf{x}_t^f と観測値 \mathbf{y}_t^o との加重平均で与えられる．

$$\mathbf{x}_t^a = \mathbf{x}_t^f + \mathbf{K}_t(\mathbf{y}_t^o - \mathbf{H}_t\mathbf{x}_t^f) \quad (3.2.6)$$

$$\mathbf{x}_t^a = (\mathbf{I} - \mathbf{K}_t\mathbf{H}_t)\mathbf{x}_t^f + \mathbf{K}_t\mathbf{y}_t^o \quad (3.2.7)$$

(3.2.6) 式の行列 \mathbf{K} は，カルマンゲインと呼ばれる重み行列である．

以上の関係式を用いると，解析誤差共分散行列 \mathbf{P}_t^a は，以下のように表せる．ただし，解析のプロセスでは，時間発展は考慮しないため，簡単のために下付きの文字 t は省略した．

$$\begin{aligned} \mathbf{P}_t^a &= \langle \mathbf{p}^a(\mathbf{p}^a)^\top \rangle \\ &= \langle (\mathbf{x}^a - \mathbf{x}^t)(\mathbf{x}^a - \mathbf{x}^t)^\top \rangle \\ &= \langle ((\mathbf{I} - \mathbf{KH})\mathbf{x}^f + \mathbf{K}\mathbf{y}^o - \mathbf{x}^t)((\mathbf{I} - \mathbf{KH})\mathbf{x}^f + \mathbf{K}\mathbf{y}^o - \mathbf{x}^t)^\top \rangle \\ &= \langle ((\mathbf{I} - \mathbf{KH})\mathbf{x}^f + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \mathbf{r}) - \mathbf{x}^t) \langle ((\mathbf{I} - \mathbf{KH})\mathbf{x}^f + \mathbf{K}(\mathbf{H}\mathbf{x}^t + \mathbf{r}) - \mathbf{x}^t)^\top \rangle \rangle \\ &= \langle (\mathbf{I} - \mathbf{KH})\mathbf{p}^f(\mathbf{p}^f)^\top(\mathbf{I} - \mathbf{KH})^\top \rangle + \langle (\mathbf{I} - \mathbf{KH})\mathbf{p}^f(\mathbf{K}\mathbf{r})^\top \rangle \\ &\quad + \langle \mathbf{K}\mathbf{r}(\mathbf{p}^f)^\top(\mathbf{I} - \mathbf{KH})^\top \rangle + \langle \mathbf{K}\mathbf{r}(\mathbf{r})^\top\mathbf{K}^\top \rangle \\ &= (\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^\top + \mathbf{K}\mathbf{R}\mathbf{K}^\top \end{aligned} \quad (3.2.8)$$

ここで，予報誤差と観測誤差は相関がなく直交していると仮定して，それらのクロスタームは0とした．最適推定値 \mathbf{x}^a を求めるためには，解析誤差分散を最小にするようなカルマンゲイン \mathbf{K} を求める必要がある．そのために，解析誤差分散と同値である \mathbf{P}^a のトレースを \mathbf{K} で偏微分することを考える．

$$\frac{\partial}{\partial \mathbf{K}}(\text{trace}\mathbf{P}^a) = 0 \quad (3.2.9)$$

(3.2.9) 式が成立するときのカルマンゲイン \mathbf{K} が，もとめる \mathbf{K} である．(3.2.9) 式を変形すると，

$$\begin{aligned} \frac{\partial}{\partial \mathbf{K}}(\text{trace}\mathbf{P}^a) &= \frac{\partial}{\partial \mathbf{K}} \left[\text{trace}((\mathbf{I} - \mathbf{KH})\mathbf{P}^f(\mathbf{I} - \mathbf{KH})^\top + \mathbf{K}\mathbf{R}\mathbf{K}^\top) \right] \\ &= \frac{\partial}{\partial \mathbf{K}} \left[\text{trace}((\mathbf{P}^f - \mathbf{P}^f\mathbf{H}^\top\mathbf{K}^\top - \mathbf{K}\mathbf{H}\mathbf{P}^f + \mathbf{K}\mathbf{H}\mathbf{P}^f\mathbf{H}^\top\mathbf{K}^\top) + \mathbf{K}\mathbf{R}\mathbf{K}^\top) \right] \end{aligned} \quad (3.2.10)$$

ここで， \mathbf{P}^f は定義から対称行列であることが言えるので，

$$\begin{aligned} \text{trace}(\mathbf{P}^f\mathbf{H}^\top\mathbf{K}^\top) &= \text{trace}(\mathbf{K}\mathbf{H}\mathbf{P}^f)^\top \\ &= \text{trace}(\mathbf{K}\mathbf{H}\mathbf{P}^f) \end{aligned}$$

が成立する．このことを用いると，(3.2.10) 式は

$$\frac{\partial}{\partial \mathbf{K}}(\text{trace}\mathbf{P}^a) = \frac{\partial}{\partial \mathbf{K}} \left[\text{trace}(\mathbf{P}^a) - 2\text{trace}(\mathbf{K}\mathbf{H}\mathbf{P}^f) + \text{trace}(\mathbf{K}\mathbf{H}\mathbf{P}^f\mathbf{H}^\top\mathbf{K}^\top) + \text{trace}(\mathbf{K}\mathbf{R}\mathbf{K}^\top) \right] \quad (3.2.11)$$

と変形できる．ここで，行列の公式

$$\frac{\partial}{\partial \mathbf{A}} \left(\text{trace}(\mathbf{A}\mathbf{B}\mathbf{A}^\top) \right) = \mathbf{A}(\mathbf{B} + \mathbf{B}^\top) \quad (3.2.12)$$

$$\frac{\partial}{\partial \mathbf{A}} (\text{trace}(\mathbf{A}\mathbf{B})) = \mathbf{B}^\top \quad (3.2.13)$$

を用いると,

$$\begin{aligned}\frac{\partial}{\partial \mathbf{K}}(\text{trace} \mathbf{P}^a) &= -2\mathbf{P}^f \mathbf{H}^T + 2\mathbf{K} \mathbf{H} \mathbf{P}^f \mathbf{H}^T + 2\mathbf{K} \mathbf{R} \\ &= -2(\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}^f \mathbf{H}^T + 2\mathbf{K} \mathbf{R} = 0\end{aligned}\quad (3.2.14)$$

という式が得られる. よって, (3.2.14) 式より求めるカルマンゲイン \mathbf{K} は,

$$\mathbf{K} = \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \quad (3.2.15)$$

となる. この式は (3.1.17) 式の第2項の係数と同じ形をしており, 予報誤差と観測誤差の和に対する予報誤差の比を表している. また, この式は以下の形に変形できる.

$$\begin{aligned}\mathbf{K} &= \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right) \mathbf{P}^f \mathbf{H}^T (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \\ &= \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} (\mathbf{H}^T + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{P}^f \mathbf{H}^T) (\mathbf{H} \mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \\ &= \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{R} + \mathbf{H} \mathbf{P}^f \mathbf{H}) (\mathbf{H} \mathbf{P}^f \mathbf{H} + \mathbf{R})^{-1} \\ &= \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1}\end{aligned}\quad (3.2.16)$$

(3.2.8) 式に (3.2.16) 式を代入すると,

$$\begin{aligned}\mathbf{P}^a &= \left[\mathbf{I} - \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right] \mathbf{P}^f \left[\mathbf{I} - \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right] \\ &\quad + \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{R} \mathbf{R}^{-1} \mathbf{H} \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-T}\end{aligned}$$

「ある2つの対称行列の積が対称行列ならば, もとの2つの対称行列は可換である」という性質を用いると, 上式は, 以下のように整理することができる.

$$\begin{aligned}\mathbf{P}^a &= \mathbf{P}^f \left[\mathbf{I} - \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \right]^T \\ &= \left[\mathbf{I} - \left((\mathbf{P}^f)^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \right] \mathbf{P}^f \\ &= (\mathbf{I} - \mathbf{K} \mathbf{H}) \mathbf{P}^f\end{aligned}\quad (3.2.17)$$

この (3.2.17) 式は予報誤差共分散行列 \mathbf{P}^f にカルマンゲイン \mathbf{K} を作用させることで, 解析誤差共分散行列 \mathbf{P}^a の各成分が小さくなることを表している.

以上で導出した (3.2.1) 式, (3.2.4) 式, (3.2.6) 式, (3.2.15) 式, (3.2.16) 式の5つの式がカルマンフィルタ KF を構成する主要な式である.

しかしながら, 導出した KF のアルゴリズムは線形性を仮定しており, このままの形では非線形のモデルに適応させることができない. そのため, 通常は何らかの近似を施してモデルに KF を組み込む. その近似を行う代表的な方法が接線形化近似である. これを行うには非線形モデル M を \mathbf{x}_0 のまわりでテーラー展開する. ただし, $\Delta \mathbf{x}$ は微小変位である.

$$M(\mathbf{x}_0 + \Delta \mathbf{x}) = M(\mathbf{x}_0) + \mathbf{M}_{\mathbf{x}_0} \Delta \mathbf{x} + O(\Delta \mathbf{x}^2) \quad (3.2.18)$$

この (3.2.18) 式の二次以上の項を無視することで線形化する. すなわち, \mathbf{M} の代わりに, M のヤコビ行列 (接線形モデル)

$$\mathbf{M}_{\mathbf{x}_0} = \left. \frac{\partial M}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \quad (3.2.19)$$

を用いることで, KF のプロセスを非線形モデルにも適応できる. この近似を行った KF を拡張カルマンフィルタ (Extended Kalman Filter: EKF) という. 一般に KF といえば, この EKF を指す場合が多い.

3.3 アンサンブルカルマンフィルタ

前節で述べた KF を実際の大気モデルに適用させようとする、計算機資源の不足という重大な問題が生じる。一般的な大気モデルの自由度は $N \sim O(10^8)$ 程度であり、解像度を粗くし、物理過程を簡略化したとしても $N \sim O(10^5)$ を下回ることはない。仮に、大気モデルの自由度が $N \sim O(10^5)$ であったとしても、誤差共分散行列の自由度は $N^2 \sim (10^{10})$ となる。さらに、それらの演算には $N^3 \sim O(10^{15})$ の計算回数が必要となる。そのため、KF を実際の大気モデルに適用させるのは不可能である。

この問題を回避する 1 つの方法として、誤差共分散行列 \mathbf{P} をアンサンブルで近似する方法が提案されている (Evensen 1994)。その手法はアンサンブルカルマンフィルタ (EnKF) と呼ばれている。以下、EnKF の導出を行う。

まず、状態ベクトルを束ねた、すなわちそれぞれの状態ベクトルを行列の列とした行列 (アンサンブル) を定義する。

$$\mathbf{X} = [\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots, \mathbf{x}^{(m)}] \quad (3.3.1)$$

ここで、 m はアンサンブルサイズである。同様に摂動ベクトル $\delta\mathbf{x}$ を束ねた行列も定義する。

$$\delta\mathbf{X} = [\delta\mathbf{x}^{(1)}, \delta\mathbf{x}^{(2)}, \delta\mathbf{x}^{(3)}, \dots, \delta\mathbf{x}^{(m)}] \quad (3.3.2)$$

ここでの摂動ベクトルは、アンサンブル平均からのずれで表現されており、摂動行列 $\delta\mathbf{X}$ は以下のように表現される。ただし、変数の上線 (バー) はアンサンブル平均を表す。

$$\delta\mathbf{X} = \mathbf{X} - \bar{\mathbf{X}} \quad (3.3.3)$$

この (3.3.3) 式を用いて、誤差共分散行列を表現すると

$$\mathbf{P} \approx \frac{\delta\mathbf{X}\delta\mathbf{X}^T}{m-1} \quad (3.3.4)$$

となる。(3.3.4) 式は、高々 m 個のメンバーで誤差共分散行列を近似することを示している。ここで、 $\mathbf{E} = \frac{\delta\mathbf{X}}{\sqrt{m-1}}$ とおくと、(3.3.4) 式は次のように書き換えることができる。

$$\mathbf{P} = \mathbf{E}\mathbf{E}^T \quad (3.3.5)$$

(3.3.5) 式より、 \mathbf{E} は誤差共分散行列の平方根と言えることがわかる。実際に、 \mathbf{P} は非負の実対称行列であるから、実行列の平方根を持つ。すなわち、(3.3.5) 式を満たす行列が存在することが正当化される。

さて、(3.3.5) 式を (3.2.4) 式に代入すると、

$$\begin{aligned} \mathbf{P}_t^f &= \mathbf{M}\mathbf{E}_{t-1}^a(\mathbf{E}_{t-1}^a)^T\mathbf{M}^T \\ &= \mathbf{M}\mathbf{E}_{t-1}^a(\mathbf{M}\mathbf{E}_{t-1}^a)^T \end{aligned} \quad (3.3.6)$$

となる。(3.3.6) 式の左辺は予報誤差共分散行列であり、この平方根を考えると、

$$\mathbf{E}_t^f = \mathbf{M}\mathbf{E}_{t-1}^a \quad (3.3.7)$$

を得る。これは誤差共分散行列 \mathbf{P} の平方根 \mathbf{E} の各列の誤差ベクトルを時間発展させるプロセスとみなすことができる。すなわち、(3.3.7) 式は行列 \mathbf{E} をアンサンブル摂動とみなしたアンサンブル予報と考えることができる。ただし、非線形モデルの下で、(3.3.7) 式は

$$\begin{aligned}\mathbf{M}\mathbf{E}_{t-1}^a &= \frac{1}{\sqrt{m-1}} \left[\mathbf{M}\delta\mathbf{x}_{t-1}^{a(1)}, \mathbf{M}\delta\mathbf{x}_{t-1}^{a(2)}, \dots, \mathbf{M}\delta\mathbf{x}_{t-1}^{a(m)} \right] \\ &\simeq \frac{1}{\sqrt{m-1}} \left[M(\bar{\mathbf{x}}_{t-1}^a + \delta\mathbf{x}_{t-1}^{a(1)}) - M(\bar{\mathbf{x}}_{t-1}^a), \dots, M(\bar{\mathbf{x}}_{t-1}^a + \delta\mathbf{x}_{t-1}^{a(m)}) - M(\bar{\mathbf{x}}_{t-1}^a) \right] \\ &\simeq \frac{1}{\sqrt{m-1}} \left[M(\mathbf{x}_{t-1}^{a(1)}) - \bar{\mathbf{x}}_t^f, \dots, M(\mathbf{x}_{t-1}^{a(m)}) - \bar{\mathbf{x}}_t^f \right]\end{aligned}\quad (3.3.8)$$

となる。

次に EnKF におけるカルマンゲイン \mathbf{K} について考える。(3.2.15) 式の \mathbf{P} に (3.3.5) 式を代入すると、

$$\begin{aligned}\mathbf{K} &= \mathbf{P}^f \mathbf{H}^T (\mathbf{H}\mathbf{P}^f \mathbf{H}^T + \mathbf{R})^{-1} \\ &= \mathbf{E}^f (\mathbf{H}\mathbf{E}^f)^T \left[\mathbf{H}\mathbf{E}^f (\mathbf{H}\mathbf{E}^f)^T + \mathbf{R} \right]^{-1}\end{aligned}\quad (3.3.9)$$

となる。このようにすることで、 $N \times N$ 行列 \mathbf{P} を直接メモリ上に保存しておく必要がなくなる。代わりに $N \times m$ 行列 \mathbf{E} と、 $ny \times m$ 行列 $\mathbf{H}\mathbf{E}$ のみを保存するだけでよい。なお、 ny は観測データの個数、すなわち観測空間の次元である。また、このようにおくことで観測演算子 H が非線形であったとしても近似計算ができるという利点も生じる。

$$\begin{aligned}\mathbf{H}\mathbf{E}^f &= \frac{1}{\sqrt{m-1}} \left[\mathbf{H}\delta\mathbf{x}^{f(1)}, \dots, \mathbf{H}\delta\mathbf{x}^{f(m)} \right] \\ &\simeq \frac{1}{\sqrt{m-1}} \left[H(\bar{\mathbf{x}}^f + \delta\mathbf{x}^{f(1)}) - H(\bar{\mathbf{x}}^f), \dots, H(\bar{\mathbf{x}}^f + \delta\mathbf{x}^{f(m)}) - H(\bar{\mathbf{x}}^f) \right]\end{aligned}\quad (3.3.10)$$

最後に、解析誤差共分散行列 \mathbf{P}^a を考える。今 \mathbf{P} はアンサンブルで表現されているため、 \mathbf{P}^a も \mathbf{P}^f と同じく解析アンサンブル摂動 \mathbf{E}^a によって計算される。そのため、このプロセスはアンサンブル・アップデートと呼ばれる。

アンサンブル・アップデートには大きく分けて2種類の方法が知られている。1つは摂動観測法 (Perturbed Observation method: PO法) と呼ばれる方法で、観測にも摂動を与え、アンサンブルメンバーを作り出す方法である。もう1つは、平方根フィルタ (Square Root Filter: SRF) と呼ばれる方法で、観測に摂動を与えずに、(3.2.17) 式を直接解くことでアンサンブル・アップデートを行う。

ここでは、PO法による EnKF の説明を行う。

まず、 \mathbf{x}^a 、 \mathbf{x}^f 、 \mathbf{y}^o をアンサンブル平均とそこからの摂動の部分に分離することを考える。

$$\mathbf{x}^a = \bar{\mathbf{x}}^a + \delta\mathbf{x}^a \quad (3.3.11)$$

$$\mathbf{x}^f = \bar{\mathbf{x}}^f + \delta\mathbf{x}^f \quad (3.3.12)$$

$$\mathbf{y}^o = \bar{\mathbf{y}}^o + \delta\mathbf{y}^o \quad (3.3.13)$$

PO法では、それぞれのアンサンブルメンバーについて独立に解析を行うので、 \mathbf{x}^a 、 \mathbf{x}^f 、 \mathbf{y}^o について (3.2.6) 式が成り立つ。すなわち、

$$\bar{\mathbf{x}}^a + \delta\mathbf{x}^a = \bar{\mathbf{x}}^f + \delta\mathbf{x}^f + \mathbf{K} \left[\bar{\mathbf{y}}^o + \delta\mathbf{y}^o - \mathbf{H}(\bar{\mathbf{x}}^f + \delta\mathbf{x}^f) \right] \quad (3.3.14)$$

を得る。ここでアンサンブル平均についても (3.3.14) 式は成立するので、

$$\delta \mathbf{x}^a = \delta \mathbf{x}^f + \mathbf{K} \left[\delta \mathbf{y}^o - \mathbf{H} \delta \mathbf{x}^f \right] \quad (3.3.15)$$

も成り立つことになる。

観測に摂動がないとき ($\delta \mathbf{y}^o = 0$) , (3.3.15) 式は

$$\delta \mathbf{x} = [\mathbf{I} - \mathbf{K}\mathbf{H}] \delta \mathbf{x}^f \quad (3.3.16)$$

と書ける。全てのアンサンブルメンバーについて (3.3.16) 式が成り立つので、以下の式を得る。

$$\delta \mathbf{X}^a = (\mathbf{I} - \mathbf{K}\mathbf{H}) \delta \mathbf{X}^f \quad (3.3.17)$$

(3.3.17) 式の両辺に $(m-1)^{-\frac{1}{2}}$ を乗じることで、

$$\mathbf{E}^a = (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{E}^f \quad (3.3.18)$$

を得る。したがって、解析誤差共分散行列 \mathbf{P}^a は、

$$\begin{aligned} \mathbf{P}^a &= \mathbf{E}^a (\mathbf{E}^a)^T \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{P}^f (\mathbf{I} - \mathbf{K}\mathbf{H}) \end{aligned} \quad (3.3.19)$$

となる。しかしながら、(3.3.19) 式は (3.2.17) 式と比較して $(\mathbf{I} - \mathbf{K}\mathbf{H})$ が一つ多く掛かっており、解析誤差を必要以上に小さく見積もってしまうことがわかる。この原因は、(3.2.8) 式と比較して、 $\mathbf{K}\mathbf{R}\mathbf{K}^T$ にあたる項がないことによる。このため、観測データに摂動を与えて、 $\mathbf{K}\mathbf{R}\mathbf{K}^T$ の項に相当するものを擬似的に作り出す必要がある。

以上が PO 法の説明である。また、EnKF の概念図は図 1 に示した。

続いて、平方根フィルタ (SRF) の説明を行う。SRF は、アンサンブル平均を最も確からしい状態とみなして、解析方程式 (3.2.6) をアンサンブル平均に対してのみ解く。つまり、解析方程式に現れる \mathbf{x} を $\bar{\mathbf{x}}$ で置き換える。これは PO 法が各メンバーについてそれぞれ解析方程式を解いていたのと対照的である。摂動成分に関しては、解析アンサンブル摂動を予報アンサンブル摂動の線形結合で与えることにより解析する。

$$\mathbf{E}^a = \mathbf{E}^f \mathbf{W} \quad (3.3.20)$$

ここで、 \mathbf{W} は結合係数からなる行列であり、解析加重行列、あるいはアンサンブル変換行列とも呼ばれる。予報アンサンブル摂動を解析アンサンブル摂動にアップデートするプロセスであることから、このプロセスをアンサンブル・アップデートともいう。(3.3.20) 式を EnKF における誤差共分散の解析方程式に代入することで、以下の式を得る。

$$\mathbf{E}^f \mathbf{W} \mathbf{W}^T (\mathbf{E}^f)^T = (\mathbf{I} - \mathbf{K}\mathbf{H}) \mathbf{E}^f (\mathbf{E}^f)^T \quad (3.3.21)$$

この式を \mathbf{W} について解くことによって、 \mathbf{W} を得る。この \mathbf{W} を使って、(3.3.20) 式の摂動成分の線形変換を実行することで、摂動成分の解析を実現する。ところで、 \mathbf{W} の選び方は一意ではなく、その平方根行列の自由度の積だけの自由度を持つことが知られている。ここでは、本研究で用いた局所アンサンブル変換カルマンフィルタについて説明する。

局所アンサンブルカルマンフィルタ (LETKF) とは、[Hunt et al. \(2007\)](#) によって提案された SRF 法の中の一つの手法である。LETKF の特徴としては、背景誤差局所化ではなく、観測

誤差局所化を行っていること、各格子点の局所領域を完全に独立に扱い、KFの解析を行うことなどが挙げられる。これらの特徴は、並列計算を行う上では有利であり、実際、Miyoshi and Yamane (2007) では、地球シミュレータに LETKF を適用し、約 99.99 % という高い並列化率を達成している。以下、LETKF のアルゴリズムについて述べる。

解析アンサンブル・アップデート (3.3.21) の右辺にカルマンゲイン (3.3.9)、共分散行列 (3.3.5) を代入して変形すると、

$$(\mathbf{I} - \mathbf{KH})\mathbf{E}^f(\mathbf{E}^f)^T = \mathbf{E}^f[(m-1)\mathbf{I} + (\mathbf{HE}^f)^T\mathbf{R}^{-1}\mathbf{HE}^f]^{-1}(\mathbf{E}^f)^T \quad (3.3.22)$$

を得る。(3.3.21) 式の左辺と比べると、

$$\mathbf{W}\mathbf{W}^T = [(m-1)\mathbf{I} + (\mathbf{HE}^f)^T\mathbf{R}^{-1}\mathbf{HE}^f]^{-1} \quad (3.3.23)$$

となる。ここで、固有値分解

$$(m-1)\mathbf{I} + (\mathbf{HE}^f)^T\mathbf{R}^{-1}\mathbf{HE}^f = \mathbf{V}\mathbf{D}\mathbf{V}^T \quad (3.3.24)$$

で固有値からなる対角行列 \mathbf{D} および固有ベクトルからなる直交行列 \mathbf{V} を定義する。(3.3.24) 式の左辺は実対称行列であるので、 \mathbf{D} および \mathbf{V} は実行列である。これを使って (3.3.23) 式を解くと、

$$\mathbf{W} = \sqrt{m-1}\mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^T \quad (3.3.25)$$

を得る。この線形変換は、(3.3.20) より予報アンサンブル摂動から解析アンサンブル摂動への写像を与える。解析アンサンブル摂動から見積もられる解析誤差共分散行列により、カルマンゲインは、

$$\mathbf{K} = \mathbf{E}^f\mathbf{V}\mathbf{D}^{-1}\mathbf{V}^T(\mathbf{HE}^f)^T\mathbf{R}^{-1} \quad (3.3.26)$$

と書ける。これを使ってアンサンブル平均の解析を行う。以上のアンサンブル平均と摂動の解析方程式をあわせて、LETKF の解析方程式

$$\begin{aligned} \mathbf{X}^a &= \bar{\mathbf{X}}^f + \mathbf{K}(\mathbf{y}^o - H(\bar{\mathbf{X}}^f)) + \delta\mathbf{X}^a \\ &= \bar{\mathbf{X}}^f + \mathbf{E}^f[\mathbf{V}\mathbf{D}^{-1}\mathbf{V}^T(\mathbf{HE}^f)^T\mathbf{R}^{-1}(\mathbf{y}^o - H(\bar{\mathbf{X}}^f)) + \sqrt{m-1}\mathbf{V}\mathbf{D}^{-1/2}\mathbf{V}^T] \end{aligned} \quad (3.3.27)$$

を得る。以上が LETKF のアルゴリズムである。

3.4 変分法

変分法とは、3.1.2 節で述べた最尤推定法を基礎とする同化手法である。すなわち、第一推定値と観測値の両方に最も近い状態を推定するために、「近さ」を端的に表す指標として評価関数を導入する。そして、最適化理論を用い評価関数を最小化させ、そのときの制御変数を解析値とする。

変分法のなかでも、誤差の時間発展を考慮しないものを 3次元変分法 (3DVAR) (図 2)、考慮するものを 4次元変分法 (4DVAR) (図 3) と言う。本研究では前者の方法を用いたので以下、3DVAR の導出を行う。

3.4.1 3次元変分法

まず、評価関数 J を多次元に拡張することを考える。第一推定値 \mathbf{x}^b およびその背景誤差（予報誤差） \mathbf{B} が与えられたとき、状態変数 \mathbf{x} の確率密度分布は次のように表せる。

$$p_b(\mathbf{x}^b|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{N}{2}}(\det\mathbf{B})^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b)\right] \quad (3.4.1)$$

同様に、観測値 \mathbf{y}^o およびその観測誤差 \mathbf{R} が与えられたときの \mathbf{x} の確率密度分布は次のように表せる。

$$p_o(\mathbf{y}^o|\mathbf{x}) = \frac{1}{(2\pi)^{\frac{ny}{2}}(\det\mathbf{R})^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}^o)^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}^o)\right] \quad (3.4.2)$$

ここで、3.1.2 節と同様に尤度関数を求めると、

$$L(\mathbf{x}|\mathbf{x}^b, \mathbf{y}^o) \propto \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) - \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}^o)^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}^o)\right] \quad (3.4.3)$$

となるので、求める評価関数 J は以下のように表せる。

$$J(\mathbf{x}) = \frac{1}{2}(\mathbf{x} - \mathbf{x}^b)^T \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \frac{1}{2}(\mathbf{H}\mathbf{x} - \mathbf{y}^o)^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}^o) \quad (3.4.4)$$

この (3.4.3) 式は、誤差分散の逆数を重みとして用いた、第1項の第一推定値への近さと第2項の観測値への近さへの重み付き平均と考えることができる。

次に、(3.4.4) 式の最小化を行う。評価関数の最小値は、通常繰り返し計算により数値的に探索する。この最小値の探索法のことを降下法と呼ぶ。一般に降下法では評価関数の値とその勾配ベクトル $\nabla J(\mathbf{x})$ を手がかりとして最小値探索を行う。(3.4.4) 式の勾配ベクトル $\nabla J(\mathbf{x})$ は次のようになる。

$$\begin{aligned} \nabla J(\mathbf{x}) &= \frac{\partial J}{\partial \mathbf{x}} \\ &= \mathbf{B}^{-1}(\mathbf{x} - \mathbf{x}^b) + \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x} - \mathbf{y}^o) \end{aligned} \quad (3.4.5)$$

最後に、解析値（最適推定値） \mathbf{x}^a を求めよう。 \mathbf{x}^a は、評価関数の勾配が0になるときの値（ $\nabla J(\mathbf{x}^a) = 0$ ）であるから、以下の関係式を得る。

$$\mathbf{B}^{-1}(\mathbf{x}^a - \mathbf{x}^b) + \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x}^a - \mathbf{y}^o) = 0 \quad (3.4.6)$$

ここで、観測演算子 H が線形であれば、

$$\mathbf{H}\mathbf{x} - \mathbf{y} = \mathbf{H}(\mathbf{x} - \mathbf{x}^b) - (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \quad (3.4.7)$$

が成り立つ。そのとき、(3.4.5) 式は \mathbf{x} の二次関数となるので解析的に \mathbf{x}^a を求めることができる。したがって、(3.4.6) 式、(3.4.7) 式より、 \mathbf{x}^a は以下のように求まる。

$$\begin{aligned} &\mathbf{B}^{-1}(\mathbf{x}^a - \mathbf{x}^b) + \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}(\mathbf{x}^a - \mathbf{x}^b) - (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b)) = 0 \\ \iff &\mathbf{B}^{-1}(\mathbf{x}^a - \mathbf{x}^b) + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}(\mathbf{x}^a - \mathbf{x}^b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) = 0 \\ \iff &(\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})(\mathbf{x}^a - \mathbf{x}^b) - \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) = 0 \\ \iff &(\mathbf{x}^a - \mathbf{x}^b) - (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) = 0 \\ \therefore &\mathbf{x}^a = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \end{aligned} \quad (3.4.8)$$

以上が 3DVAR のアルゴリズムである。また、3DVAR の概念図は図 2 に示した。

3.4.2 前処理

一般に、大気モデルで用いられるような \mathbf{B} は、非常に大きな自由度を持っており、その逆行列 \mathbf{B}^{-1} の計算は困難である。そもそも \mathbf{B}^{-1} が存在するかどうか不明である。しかしながら、 \mathbf{B}^{-1} が存在しなければ、(3.4.12) 式を計算することはできない。その対処法の一つが前処理である。

最初に、前処理行列と呼ばれる背景誤差共分散行列の平方根 \mathbf{U} を導入する。

$$\mathbf{B} = \mathbf{U}\mathbf{U}^T \quad (3.4.9)$$

次に、以下の式を満たすような、新しい制御変数 \mathbf{s} を導入する。

$$\mathbf{x}^a = \mathbf{x}^b + \mathbf{U}\mathbf{s} \quad (3.4.10)$$

(3.4.2) 式の \mathbf{s} を用いて、評価関数を表現すると以下のようになる。

$$J(\mathbf{s}) = \frac{1}{2}\mathbf{s}^T\mathbf{s} + \frac{1}{2}(\mathbf{H}\mathbf{U}\mathbf{s} + \mathbf{v})^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\mathbf{s} + \mathbf{v}) \quad (3.4.11)$$

ただし、 \mathbf{v} はイノベーションと呼ばれるベクトルで、以下のようにおいた。

$$\mathbf{v} = H(\mathbf{x}) - \mathbf{y}^o \quad (3.4.12)$$

また、評価関数の勾配は以下のようになる。

$$\nabla_{\mathbf{s}}J = \mathbf{s} + \mathbf{U}^T\mathbf{H}^T\mathbf{R}^{-1}(\mathbf{H}\mathbf{U}\mathbf{s} + \mathbf{v}) \quad (3.4.13)$$

以上が前処理を行った 3DVAR である。

3.5 ハイブリッドデータ同化

ハイブリッドデータ同化とは、Hamil and Snyder (2000) によって初めて考案された手法であり、他にも Lorenc (2003), Wang et al. (2008) など様々な先行研究がある。ここでは、本研究で導入した Hamil and Snyder (2000) によって考案された一番簡単な形式の Hybrid-3DVAR について述べる。

(3.4.4) 式で示した評価関数の式に登場する背景誤差共分散行列を、アンサンブル予報による背景誤差共分散行列（予報誤差共分散行列）によってある程度置き換える、すなわち元々の気候値的な背景誤差共分散とアンサンブルによる流れ依存性を持った共分散の加重平均のようなものを考える。

$$\mathbf{B} = \alpha_1\mathbf{B}_1 + \alpha_2\mathbf{B}_2 \quad (3.5.1)$$

ただし、 $\alpha_1 + \alpha_2 = 1$ である。また添え字の 1 は、気候値的な背景誤差共分散、添え字の 2 は、アンサンブルによる共分散を示している。この (3.5.1) 式で得られた背景誤差の情報を 3DVAR のアルゴリズムに取り込むことによって Hybrid-3DVAR を実現する。また、ハイブリッドデータ同化の模式図に関しては、図 4 に示した。

3.6 安定化

この節では、今まで説明してきた手法の安定化について説明する。

3.6.1 共分散膨張

3.3節で説明した EnKF をそのまま非線形モデルに適用させると、モデルの不完全性や線形化の影響、サンプリングエラーなどによって、誤差共分散が必要以上に小さくなる。このことをサイクルプロセスで繰り返してしまうと、観測の情報がほとんど取り込まれなくなる。そして、解析値が現実の値と乖離していき、最終的には発散してしまうことが知られている。この現象を回避するには、小さく見積もられた共分散の値を大きくしてやれば良い。具体的には δ を小さな正の値として、(3.3.6) 式を次のように書き換える。

$$\mathbf{P}_t^f = (1 + \delta) \mathbf{M} \mathbf{E}_{t-1}^a (\mathbf{M} \mathbf{E}_{t-1}^a)^T \quad (3.6.1)$$

このような (3.6.1) 式の操作を共分散膨張 (covariance inflation) と呼ぶ。また、共分散そのものを膨張させるのではなく、アンサンブル摂動を膨張させる方法もある。すなわち、(3.3.6) 式を次のように書き換えれば良い。

$$\begin{aligned} \mathbf{P}_t^f &= \mathbf{M} (1 + \delta) \mathbf{E}_{t-1}^a (\mathbf{M} (1 + \delta) \mathbf{E}_{t-1}^a)^T \\ &= (1 + \delta)^2 \mathbf{M} \mathbf{E}_{t-1}^a (\mathbf{M} \mathbf{E}_{t-1}^a)^T \end{aligned} \quad (3.6.2)$$

これらの方法は積によって共分散を膨張させているので、multiplicative covariance inflation と呼ばれる。

他の方法として、解析アンサンブルに解析誤差共分散行列を反映するようなランダム摂動 \mathbf{J} を加える方法もある。この場合は、(3.3.7) 式を以下のように書き換える。

$$\begin{aligned} \mathbf{E}_t^f &= \mathbf{M} (\mathbf{E}_{t-1}^a + \mathbf{J}) \\ &= \mathbf{M} \mathbf{E}_{t-1}^a + \mathbf{M} \mathbf{J} \end{aligned} \quad (3.6.3)$$

(3.6.3) 式からわかるように、この方法ではモデルによる変形を受けた $\mathbf{M} \mathbf{J}$ を用いることで、日々変化する誤差成分を取り出すことを可能にしている。この方法は先の積をとる手法に対して、和をとることから additive covariance inflation と呼ばれる。

3.6.2 局所化

一般に、離れた点どうしの誤差相関は小さいと考えられるので、離れた点どうしで検出された大きなシグナルには、サンプリングエラーが混入していると推測できる。このため、サンプリングエラーを小さくするために、2点間の距離が離れるほど0に近づくような関数を誤差共分散に掛けて、離れた点の相関をダンプする。このようにして、アンサンブルデータ同化におけるメンバー数の制限によって生じるサンプリングエラーを除くことを局所化という。

3.7 No-cost Ensemble Kalman Smoother と RIP スキーム

No-cost Ensemble Kalman Smoother (No-cost EnKS) とは、その名の通り、コストをかけずにアンサンブルカルマン smoother を実現する手法のことである (Yang et al. 2009)。そのアルゴリズムは以下の簡潔な式で表される。

$$\tilde{\mathbf{x}}_{n-1}^a = \delta \mathbf{X}_{n-1}^a \tilde{\mathbf{w}}_n \quad (3.7.1)$$

$$\delta\tilde{\mathbf{X}}_{n-1}^a = \delta\mathbf{X}_{n-1}^a \mathbf{W}_n \quad (3.7.2)$$

ただし、添え字の n は時間ステップ数を表す。これらの式の意味するところは、現在時刻の時間ステップより1ステップ前の解析アンサンブル摂動行列に解析加重行列を作用させることによって、簡易的にはあるが平滑化された解析アンサンブル摂動 $\delta\tilde{\mathbf{X}}_{n-1}^a$ を作成するというものである。これはいわばアンサンブルの作り直しをしているようなものと考えて良い。

続いて、RIP スキーム (Kalnay and Yang 2010; Yang et al. 2012) の説明を行う。RIP スキームとは、上述の No-cost EnKS を有効的に活用しているスキームであり、主に EnKF のスピニングの早くする目的で開発された。基本的には、上述の No-cost EnKS のアルゴリズムと変わらないが、以下に RIP スキームの説明を述べる。

まず、通常の LETKF の解析を行い、時刻 n における解析加重行列 \mathbf{W}_n を導出する。また、同時に平均二乗された観測値と観測空間に焼き直した状態ベクトルの差分を以下の式に従って算出する。

$$OMF^2 = [\mathbf{y}_n^o - H_n(\mathbf{x})]^T (\mathbf{R}_n)^{-1} [\mathbf{y}_n^o - H_n(\mathbf{x})] \quad (3.7.3)$$

次に、先程の No-cost EnKS を用いて、平滑化された解析アンサンブル摂動を計算する。その後、もともとの解析アンサンブル摂動を用いたときと同様のアンサンブル予報の結果になることを防ぐため、あるいはアンサンブル摂動をより早く成長する方向へ発達させるために平滑化された解析アンサンブル摂動に小さなガウシアンノイズを与える。これらの処理が完了したら、平滑化された解析アンサンブルを元にアンサンブル予報を行い、 OMF^2 について以下の判定を行う。

$$\frac{OMF^2(iter) - OMF^2(iter + 1)}{OMF^2(iter)} > \epsilon \quad (3.7.4)$$

式の ϵ は、適当なしきい値である。この式が真であるときは同様の計算をもう一度繰り返し、偽であるときは次の同化ウインドウに移る。また、その判定式が5回連続で偽になったときは RIP スキームを終了させ、通常の LETKF のアルゴリズムに戻す。

以上が RIP スキームの説明である。

4 実験設定

この章では、実験に用いた簡易的 AGCM である SPEEDY (Simplified Parameterizations primitive-Equation Dynamics) の説明および本研究における実験の設定を簡単に説明する。

4.1 SPEEDY

SPEEDY とは、Molteni (2003) によって開発されたスペクトルモデルであり、以下のような特徴を持っている。

- 1) 解像度は、T30L7 (格子点数は $96 \times 48 \times 7$) である。鉛直層の座標系は σ 座標系を採用しており、第 1 層から順に、1.000, 0.950, 0.835, 0.685, 0.510, 0.340, 0.200, 0.080 である。
- 2) 似たような解像度の AGCM と比較して、計算コストが 1 オーダーぐらい小さい。
- 3) 大気の大循環の挙動を比較的良く再現しているが、他の AGCM に比べると誤差の絶対値が大きくなる傾向を持っている。

本研究では、SPEEDY-LETKF および SPEEDY-3DVAR について実験を行った。以下にそれぞれの実験設定の概要を記す。

4.1.1 実験 1 LETKF, No-cost EnKS および RIP スキームの比較実験

この実験では、SPEEDY-LETKF に実装した No-cost EnKS および RIP スキームの精度比較を行った。

- 真値を 1981 年 1 月 1 日の静止大気から 1 年間のスピナップを行った後の 2 ヶ月間 (1982 年 1 月 1 日から 1982 年 3 月 1 日) のシングルラン (ネイチャーラン) とした。
- 予報変数は、水平風 (u, v)、気温 (T)、比湿 (q)、地表面気圧 (P_s) の 5 つである。
- 本実験における観測値はネイチャーランの予報変数にガウシアンノイズを加えることで作成した。観測誤差の標準偏差の設定はそれぞれ、1.0 m/s, 1.0 K, 0.0001 kg/kg, 100.0 Pa である。
- 観測地点は、Miyoshi (2005) の実験設定を参考にし、現実の観測網を模した 416 地点とした (図 5)。基本的には、観測地点の多くを陸上に作り、海上での観測はほとんどないと仮定している。
- LETKF の同化サイクルは 6 時間ごととし、解析時刻は 1982 年 1 月 1 日から 1982 年 2 月 1 日の 1 ヶ月間とした。
- 観測変数は上で述べた予報変数と同じ 5 つの変数を使用した。また、基本的には観測は地表面気圧以外は各層に与えた。
- 初期アンサンブルメンバーは同化時刻に近い月でのランダムな時刻のネイチャーランから抽出したものとした。

- アンサンブルメンバー数に関しては，大嶋（2011）を参考にし 20 とした．
- 局所化スケールに関しては，Miyoshi and Yamane (2007) を参考にし，水平方向のスケールは 1000km，鉛直方向のスケールは $0.1 \ln P$ とした．
- 共分散膨張に関しては，multiplicative covariance inflation を採用し，そのときの共分散膨張パラメータは， $\delta = 0.05$ （固定）とした．

4.1.2 実験 2 3DVAR および Hybrid-3DVAR の比較実験

この実験では，導入した Hybrid-3DVAR と 3DVAR との精度比較を行った．

- 真値および観測値の設定に関しては実験 1 に準ずる．
- 背景誤差共分散に関しては，Miyoshi (2005) の設定に準じ，1982 年 1 月 10 日から 1 月 31 日の 87 サンプルを元に NMC 法（Parrish and Derber 1992; Kalnay 2003）を適用し作成した．
- SPEEDY-3DVAR で用いた背景誤差共分散行列は前処理を適用している．その前処理行列は $\mathbf{U} = \mathbf{VCA}$ と表され，それぞれ \mathbf{V} は地衡風バランスなどの状態変数間での誤差相関， \mathbf{C} は水平方向および鉛直方向の誤差相関， \mathbf{A} は誤差の標準偏差を成分に持つ行列である．
- Hybrid-3DVAR に関しては，第 1 層から第 6 層までの背景誤差を NMC 法によるものと，アンサンブルによる背景誤差の情報を $\alpha_1 = \alpha_2 = 0.5$ の割合で重み付け平均を取った．ただし，第 7 層に関しては，アンサンブルによる背景誤差の値が発散してしまうことがあったため，NMC 法による背景誤差の重みを 1 としている．

5 結果

5.1 実験1の結果

図6は、鉛直第4層における気温のRMSEの時系列を示している。この図を見ると、どの手法もRMSEの値が順調に減少しているが、通常のLETKFに関しては、1月15日あたりで誤差が再び悪化していることがわかる。一方で、3回の繰り返しのNo-cost EnKSを除けば、その時点におけるRMSEの値は上昇していないことも見て取れる。さらに、No-cost EnKSの結果のみを比較すると、繰り返し回数が増加していくたびに、最終的なRMSEの値が悪化していることがわかる。また、RIPスキームの結果はこの中では比較的良好であった1回の繰り返しのNo-cost EnKSの精度とほぼ同等であることもわかったが、一方で最初のステップで、誤差が大きく増大してしまうこともありうることが示された。今回はアンサンブルの作り直しという点に注目したいため、以降No-cost EnKSの結果に着目していく。

図7から図10は他の出力変数に関しても同等の精度比較を行ったものである。ただし、図10に関しては6時間降水量について解析しているため、RMSEではなくて相関係数を見ている。これらの図を参照すると、水平風、比湿、地表面気圧、6時間降水量に関しても、1回の繰り返しのNo-cost EnKSの結果の精度が最も良く、繰り返しの回数を重ねていくたびに、精度が悪化しているという図6と同様の結果が得られた。

次に、通常のLETKFの解析で誤差が悪化してしまった時刻に着目し、1982年1月15日12時（世界標準時）について全球におけるRMSEおよびSpreadを計算した。図11は、通常のLETKFにおける気温のRMSE、図12から図14はNo-cost EnKSの気温のRMSEを示している。これらの図を参照すると、全球的には誤差は小さくなっていることがわかるが、南半球のチリ沖近くの部分において、比較的RMSEが大きくなっていることがわかる。特に図11および図14で顕著である。Spreadに関しても、図15から図18を見てわかる通り、同様の地域でSpreadが大きくなっていることがわかる。特に図16を見ると、他の図17や図18の結果とくらべても、同様の地域でSpreadが顕著に増大している。また、図19から図21は各実験結果のSpreadの差分を取ったものである。各図を見ると、全体的に暖色になっており、Spreadが減少していることがわかる。ただ、前述の南半球のチリ沖付近の領域では、南極大陸にかけて濃い寒色系、すなわちSpreadがかなり大きくなっていることがわかる。また、太平洋の低緯度域において、No-cost EnKSの繰り返し回数が増大するほどSpreadが増大傾向にあることがわかる。

5.2 実験2の結果

図22、図23はNMC法で作成した鉛直第4層における東西風および気温の背景誤差の標準偏差を示している。これらの図より、基本的に大陸で標準偏差が大きく、特に北米大陸と南米大陸で標準偏差が大きくなっている傾向が見受けられた。図24、図25は、実験1で注目した1982年1月15日12時における解析アンサンブルを用いて作成した背景誤差の標準偏差を示している。両図を見てみると、上述の南半球チリ沖付近で、特に標準偏差が大きくなっていることがわかる。これは、図15で見たような、その領域における不確実性を反映していると考えられる。

図 26, 図 27 は図 6 のように各同化手法に関して精度比較を行っているものである。図 26 を見ると、通常の 3DVAR (水色の円) が最も誤差が大きく、続いてアンサンブル予報による背景誤差のみを利用した 3DVAR (赤色の円), 続いて Hybrid-3DVAR (青色の円, 橙色の四角形, 紫色の四角形) の順に誤差が減少していることがわかる。このことから、元の 3DVAR や背景誤差の情報をすべてアンサンブルのものに置き換えた 3DVAR よりも背景誤差を混ぜ合わせた Hybrid-3DVAR のほうが精度が良いことが分かった。しかしながら、アンサンブルの情報がより正しいと期待される No-cost EnKS の解析アンサンブルを用いたもの (橙色の四角形), アンサンブルメンバー数を他の実験の 4 倍にしたもの (紫色の四角形) の精度を見てみても、通常の Hybrid-3DVAR と比較してほとんど差がないという結果となった。

また、今回の実験では、全体的に 3DVAR の精度は LETKF に及ばないことがわかった。

6 考察

6.1 実験1の考察

まず、No-cost EnKSの精度が繰り返し回数を増やす度に悪化していく要因を考察する。これは式(3.7.2)に示されている通り、1ステップ先の解析加重行列を解析アンサンブルに繰り返し作用させていることが大きな原因だと考えられる。これは、すなわちSPEEDYというモデルでは非線形性がそこまで強くないため、未来の観測情報を持つ解析加重行列を何回も作用させることで解析アンサンブルが未来の状態に寄りすぎてしまい、あるいは解析Spreadが小さくなりすぎてしまい、現在の時刻から見ると真値から大きく離れてしまったのだと推測できる。これは機械学習で言うところの過学習と同等のことが起きていると考えられる。

また、RIPスキームが最初のステップのところで大きく誤差が大きくなってしまっている要因はこれと同様と考えられる。なぜならば、RIPスキームは基本的に(3.7.4)式を満たさなければ何回でもNo-cost EnKSを繰り返し行うアルゴリズムになっているためである。実際に、最初に誤差が大きくなってしまったステップにおける繰り返し回数を調べてみたところ、繰り返し回数が10回となっていた。今回の実験結果と合わせて考えてみると、繰り返し回数が本来は1回で十分なところに10回も繰り返しを行ってしまったことで、解析アンサンブルが未来の状態に近づきすぎてしまい誤差が他の手法にくらべて大きくなってしまったのだと考えられる。しかしながら、その後のRMSEの時系列をみると、最も精度の良かった繰り返し回数が1回のNo-cost EnKSと同等の精度を見せているので、SPEEDY-LETKFの枠組みでは、最初のほうの時間ステップで精度が多少悪くならうとも学習を繰り返しておくことで、後の解析値の誤差を減らすことができる可能性があることを示唆しているとも考えられる。

次に、全球でRMSEおよびSpreadの解析を行った結果について考察する。1982年1月15日12時の時点におけるNo-cost EnKSの精度が通常のLETKFに比べて良かった要因としては、図15から図18で見たように誤差が大きくなっている南半球のチリ沖・南極大陸付近の領域で、Spreadが増大していること、すなわち不確実性が大きい領域において十分大きなSpreadを確保できていたことが考えられる。一般にEnKFの解析において、Spreadの大きさとRMSEの大きさが同程度であることが望ましい(近藤2009)ので、この結果はそのことを反映していると考えられる。しかしながら、図19で見た通り、基本的にはNo-cost EnKSのアルゴリズムでは解析加重行列を解析アンサンブルに作用させているので、Spreadは減少するはずである。すなわち、No-cost EnKSは不確実性の大きい領域のSpreadを増加させる作用もあることが示唆される。ただ、特にRMSEが大きくなっていない太平洋の低緯度域、特に北米・南米大陸付近の海域で、繰り返し回数が2回、3回のNo-cost EnKSにおいてはこのSpreadが大きくなるということは不利に働いていると考えられる。

以上をまとめると、不確実性の大きい領域が存在したときには、No-cost EnKSはその領域のSpreadを増大させ、アンサンブル予報の精度を向上させる可能性を持つが、繰り返し回数を重ねることで、不確実性が大きくない領域に関してもSpreadを大きくしてしまい、むしろアンサンブル予報の精度を悪化させてしまう可能性も持ちうることを示された。

6.2 実験2の考察

まず、背景誤差の標準偏差の図について述べていく。ここでは、図 24, 25 に現れるような不確実性の大きい領域が現れたときにも解析アンサンブルによる背景誤差を使用することで、3DVAR の精度改善に繋がることが示された。

図 26 を見てみても、1 月 15 日付近では LETKF の結果（緑色の円）と同様に、通常の 3DVAR の結果（水色の円）が精度が悪化している時刻で、Hybrid-3DVAR はそれと比較して精度が向上し得ることがわかる。しかしながら、結果の章でも述べたように、精度がより向上しているはずの解析アンサンブル（No-cost EnKS によるもの、およびメンバー数を 80 にして LETKF の解析を行ったもの）を用いた Hybrid-3DVAR と通常の Hybrid-3DVAR での精度に大きな違いは見受けられなかった。このことは、SPEEDY-Hybrid-3DVAR においては、アンサンブル予報の精度をいくら向上させても、3DVAR の精度を向上させるのには限界があることを示唆している。

7 結論

本研究では、ハイブリッドデータ同化手法に、アンサンブルの作り直しを行う手法、No-cost EnKS, RIP スキームを組み合わせたときの影響について調査した。

その結果、SPEEDY-LETKFにおいては、No-cost EnKSを実装することで、不確実性の大きい領域が現れた際に精度が良くなる可能性を持つことが示された。一方で、不確実性が大きくない領域に関しても Spread が必要以上に大きくなってしまい、精度をむしろ悪化させてしまう可能性があることも示された。

また、SPEEDY-Hybrid-3DVAR に関しては、通常の 3DVAR よりも精度が向上することが示されたが、アンサンブルの作り直しの手法 (No-cost EnKS) と組み合わせても通常の Hybrid-3DVAR の精度と大きく変わることはなかった。この結果はいくら 3DVAR により流れ依存性を持った背景誤差の情報を導入しても精度改善には限度があることを示唆している。

今後の課題としては、Hybrid-3DVAR のみでなく Hybrid-4DVAR との比較を行うこと、精度が悪化したときのみ RIP スキームを用いることができるような改良された RIP スキームの考案を試みることなどが挙げられる。

謝辞

本研究を進めるにあたり、指導教員である筑波大学計算科学研究センターの田中博教授には、他の研究室から移って来たのにもかかわらず、筆者の拙い研究結果に対する考察などへの丁寧な御指導、御助言を賜りました。心より感謝申し上げます。

筑波大学生命環境科学研究科の植田宏昭教授，日下博幸教授，上野健一准教授，釜江陽一助教，松枝未遠助教には，大気分野ゼミ，集中ゼミ，研究発表会などの発表の場を通じ，研究内容や手法，課題に関して様々な御助言を頂きました。また，気象庁気象研究所の梶野瑞王連携准教授には副指導教員を引き受けて頂きました。心より感謝申し上げます。

また，博士前期課程1年の夏期休暇の際に行った，理化学研究所計算科学研究機構におけるスチューデントインターンは私の人生の中で最も研究が楽しかった2ヶ月間であり，そのときの経験がこの修士論文の糧となりました。データ同化研究チームの皆さんそして三好建正チームリーダーに，心より感謝申し上げます。

最後になりますが，研究に関することについてはもちろん様々な場面で意見を交換し交流を深めてきた大気大循環研究室の皆様や，共に修士研究を進めてきた同期の皆様をはじめ，これまで支えてくださった全ての皆様の御協力により本論文を完成させることが出来ました。篤く御礼申し上げます。

参考文献

- [1] 淡路敏之, 蒲地政文, 池田元美, 石川洋一 編著, 2009 : データ同化—観測・実験とモデルを融合するイノベーション. 京都大学学術出版会, 284pp.
- [2] Bouttier, F. and P. Courtier, 1999 : Data assimilation concepts and methods. *Meteorological training course lecture series*, ECMWF, 75pp.
- [3] 堂山俊貴, 2014 : 浅水モデルを用いた変分法データ同化システムの開発と非線形効果の調査. 筑波大学生命環境学群地球学類卒業論文.
- [4] Evensen, G., 1994 : Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99**(C5), 10143-10162.
- [5] Hunt, B. R., E. J. Kostelich, I. Szunyogh, 2007 : Efficient data assimilation for spatiotemporal chaos: A local transform Kalman filter. *Physica D: Nonlinear Phenomena*, **230**(1), 112-126.
- [6] Hamill T. M. and C. Snyder, 2000 : A hybrid ensemble Kalman filter-3D variational analysis scheme, *Monthly Weather Review*, **128**, 2905-2919.
- [7] Kalman, R. E. 1960 : A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, **82**(1), 35-45.
- [8] Kalnay, E., 2003 : Atmospheric modeling, data assimilation and predictability, Cambridge University Press, 341pp.
- [9] Kalnay, E., Li, H., Miyoshi, T., Yang, S. C., and Ballabrera-Poy, 2007 : 4D-Var or Ensemble Kalman Filter? *Tellus* **59A**, 758-773.
- [10] Kalnay, E., and S. C. Yang, 2010 : Accelerating the spin-up of Ensemble Kalman Filtering. *Quarterly Journal of the Royal Meteorological Society*, **136**, 1644-1651.
- [11] 近藤圭一, 2009 : NICAM-LETKF の開発および4次元データ同化実験. 筑波大学大学院生命環境科学研究科地球科学専攻修士論文.
- [12] Lorenc, A. C., 2003 : The potential of the ensemble Kalman filter for NWP—A comparison with 4D-VAR. *Quarterly Journal of the Royal Meteorological Society*, **129**, 3183-3203.
- [13] Miyoshi, T., 2005 : Ensemble Kalman filter experiments with a primitive-equation global model. Ph. D. dissertation, University of Maryland, 197pp.
- [14] 三好建正, 2005 : アンサンブル・カルマンフィルタ—データ同化とアンサンブル予報の接点—. *天気*, **52**, 93-104.
- [15] 三好建正, 2006 : アンサンブル・カルマンフィルタ～データ同化との融合～. 数値予報課報告・別冊第52号, 80-99.

- [16] Miyoshi, T. and S. Yamane, 2007 : Local ensemble transform Kalman filtering with an AGCM at a T159/L48 resolution. *Monthly Weather Review*, **135**, 3841-386.
- [17] Molteni, F. 2003 : Atmospheric simulations using a GCM with simplified physical parametrizations. I: Model climatology and variability in multi-decadal experiments. *Climate Dynamics*, **20**(2-3), 175-191.
- [18] 大嶋 彬, 2011 : 予測可能性から見た Local Ensemble Transform Kalman Filter の性能調査 -大気大循環モデルを用いて-. 北海道大学大学院環境科学院地球圏科学専攻大気海洋物理学・気候力学コース修士論文.
- [19] Parrish, D. and J. Derber, 1992 : The National Meteorological Center ' s spectral statistical interpolation analysis system. *Monthly Weather Review*, **120**, 1747-1763.
- [20] 露木 義・川畑拓矢 編, 2008 : 気象学におけるデータ同化. 気象研究ノート, **217**, 277pp.
- [21] Yang, S. C., Corazza, M., Carrassi, A., Kalnay, E., and Miyoshi, T. 2009 : Comparison of ensemble-based and variational-based data assimilation schemes in a quasi-geostrophic model. *Monthly Weather Review*, **137**, 639-709.
- [22] Yang, S. C., E. Kalnay, B. Hunt, 2012: Handling Nonlinearity in an Ensemble Kalman Filter: Experiments with the Three-Variable Lorenz Model. *Monthly Weather Review*, **140**, 2628-2646.
- [23] Wang, X., D. Barker, C. Snyder, T. M. Hamill, 2008 : A hybrid ETKF-3DVAR data assimilation scheme for the WRF model. Part I: observing system simulation experiment. *Monthly Weather Review*, **136**, 5116-5131.

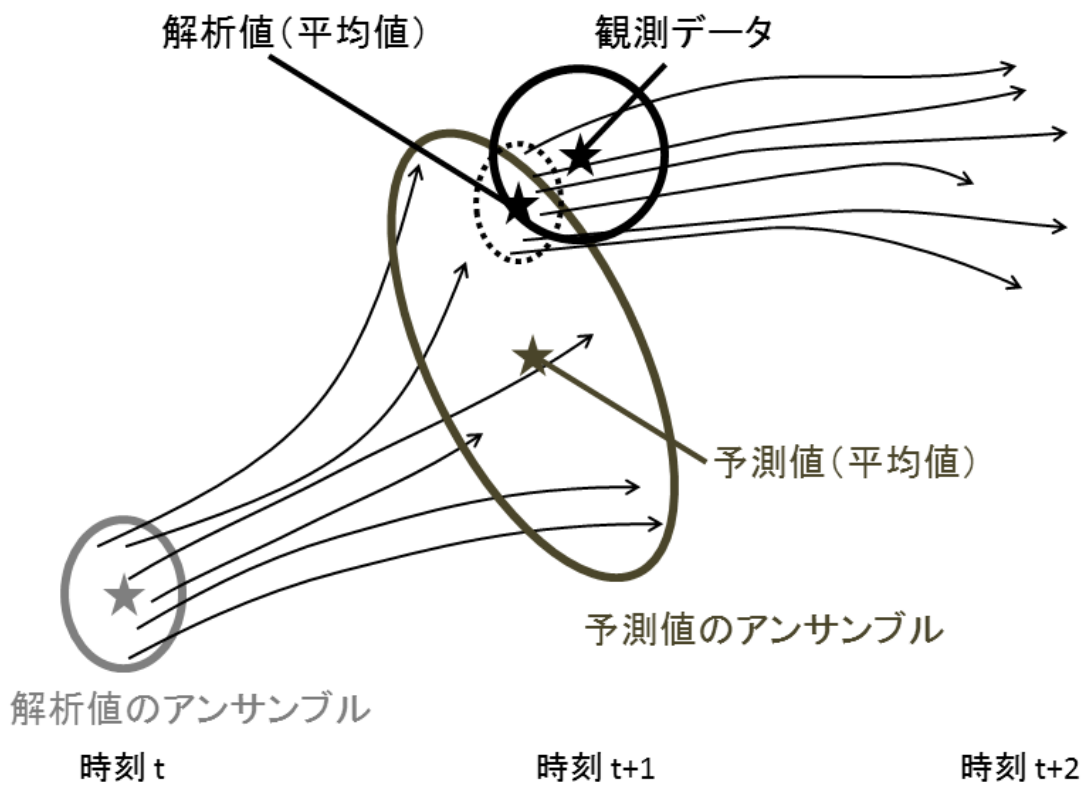


図 1: アンサンブルカルマンフィルタ (EnKF) の概念図.

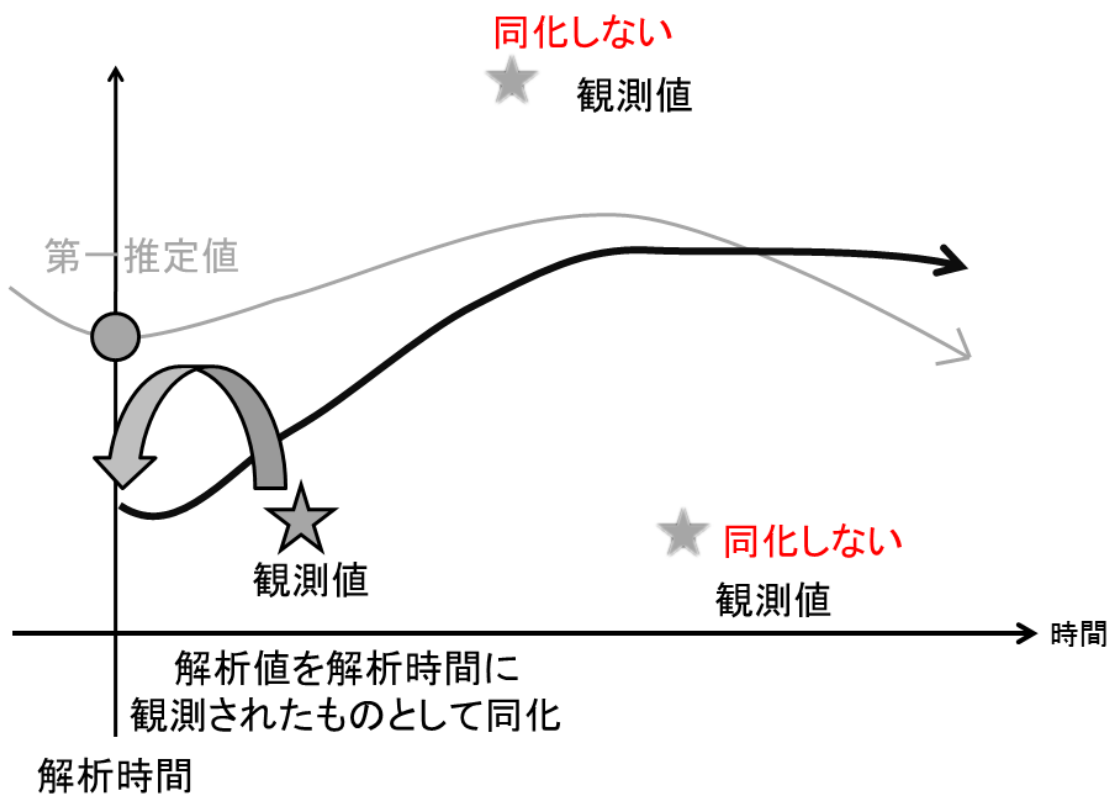


図 2: 3次元変分法 (3DVAR) の概念図. 灰色の矢印が同化を行う前の予報値のトラジェクトリを表している.

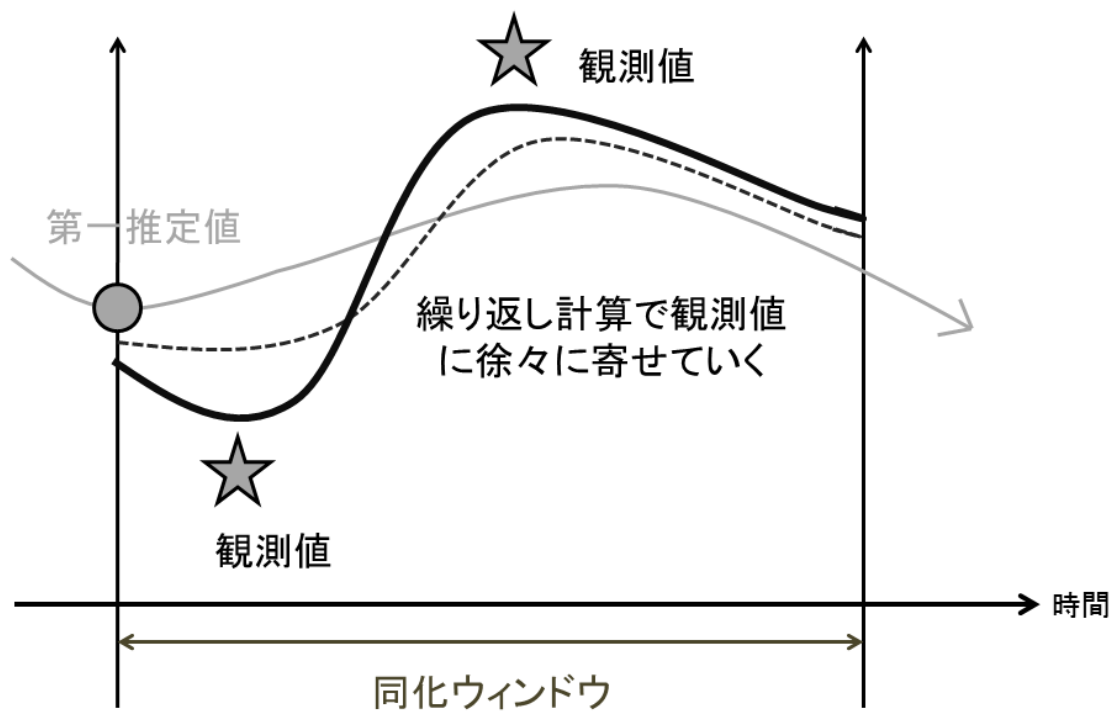


図 3: 4次元変分法（4DVAR）の概念図．灰色の矢印が同化を行う前の予報値のトラジェクトリを表している．

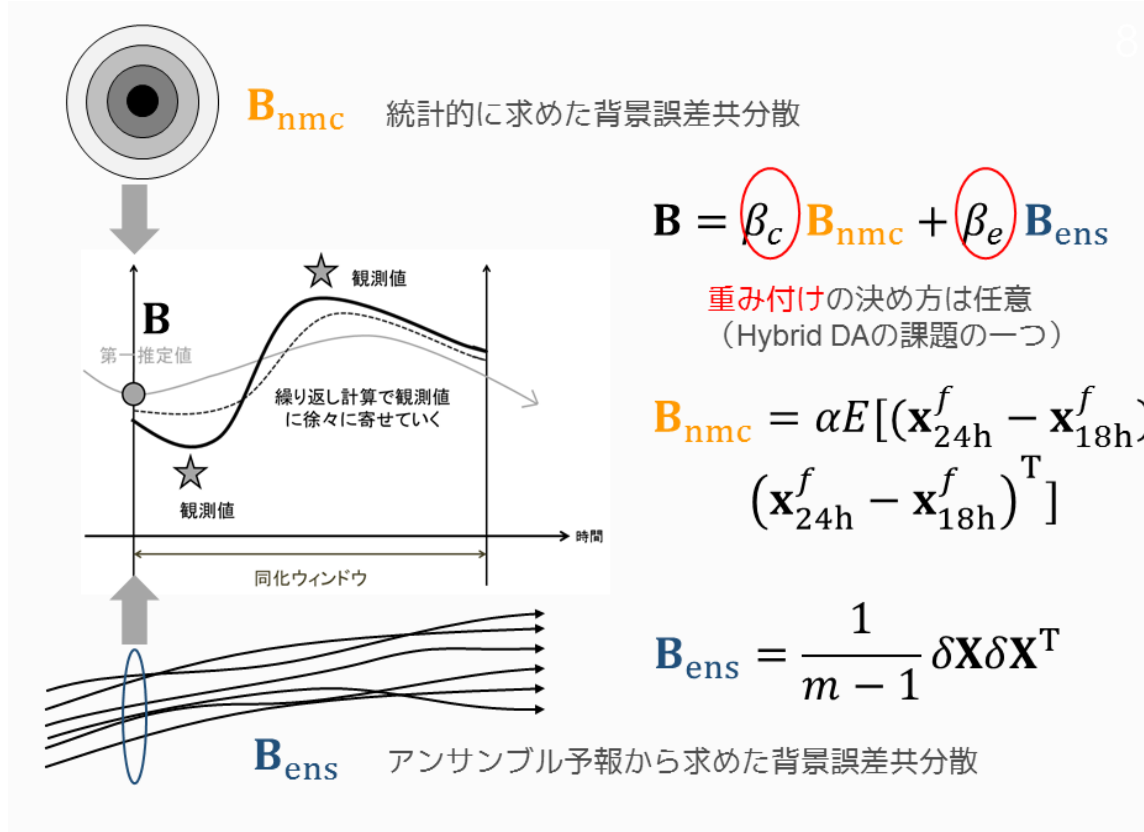


図 4: ハイブリッドデータ同化の概念図. **B_{nmc}** は NMC 法による統計的な背景誤差共分散, **B_{ens}** はアンサンブルから求めた流れ依存性を持つ背景誤差共分散を表す.

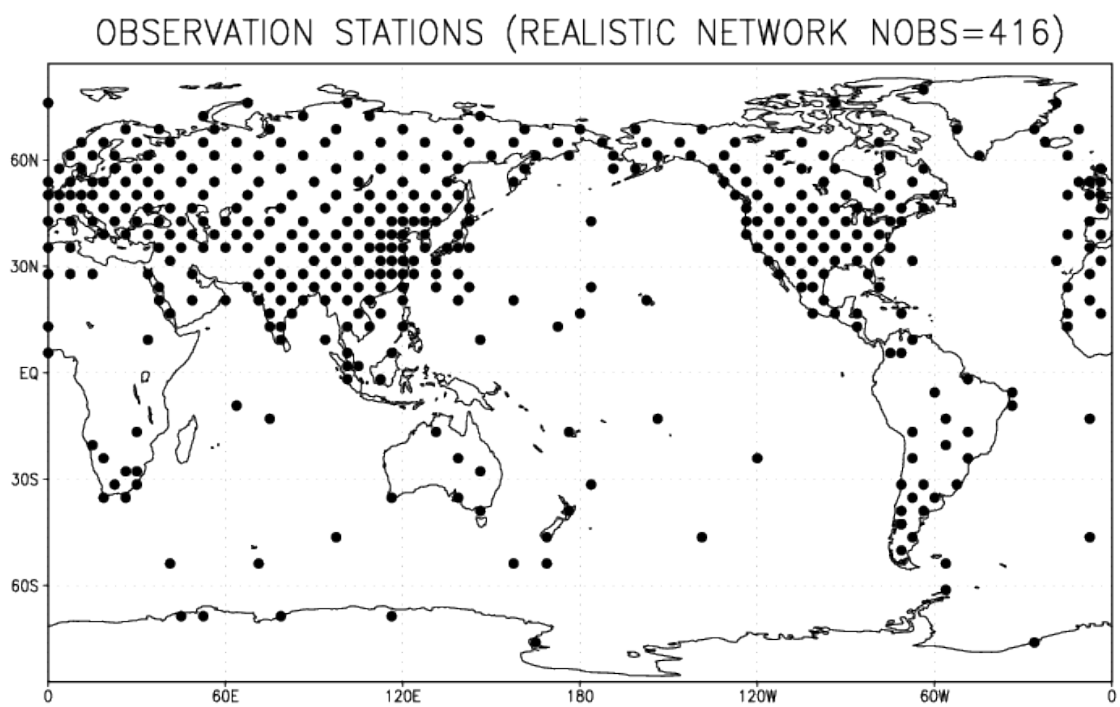


図 5: 本実験で用いた観測分布. 黒色の点が観測地点を表している. 観測密度は9.03%となっている.

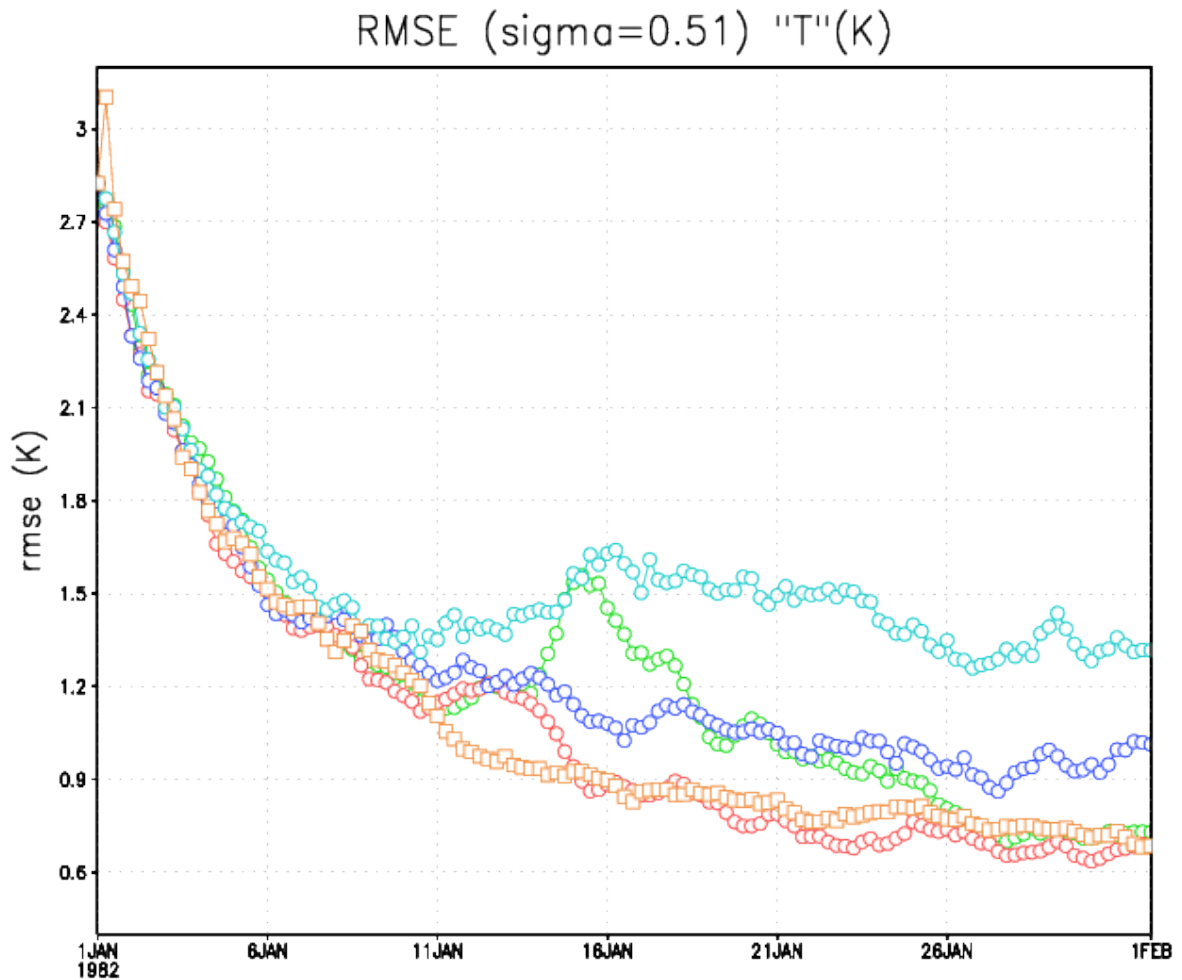


図 6: 鉛直第 4 層における気温の解析二乗平均平方根誤差 (RMSE)。縦軸が RMSE の値で、横軸が時系列を表している。それぞれ、緑色の円が通常の LETKF、赤色の円が 1 回の繰り返し No-cost EnKS、青色の円が 2 回の繰り返し No-cost EnKS、水色の円が 3 回の繰り返し No-cost EnKS、オレンジ色の四角形が RIP スキームによる結果を示している。

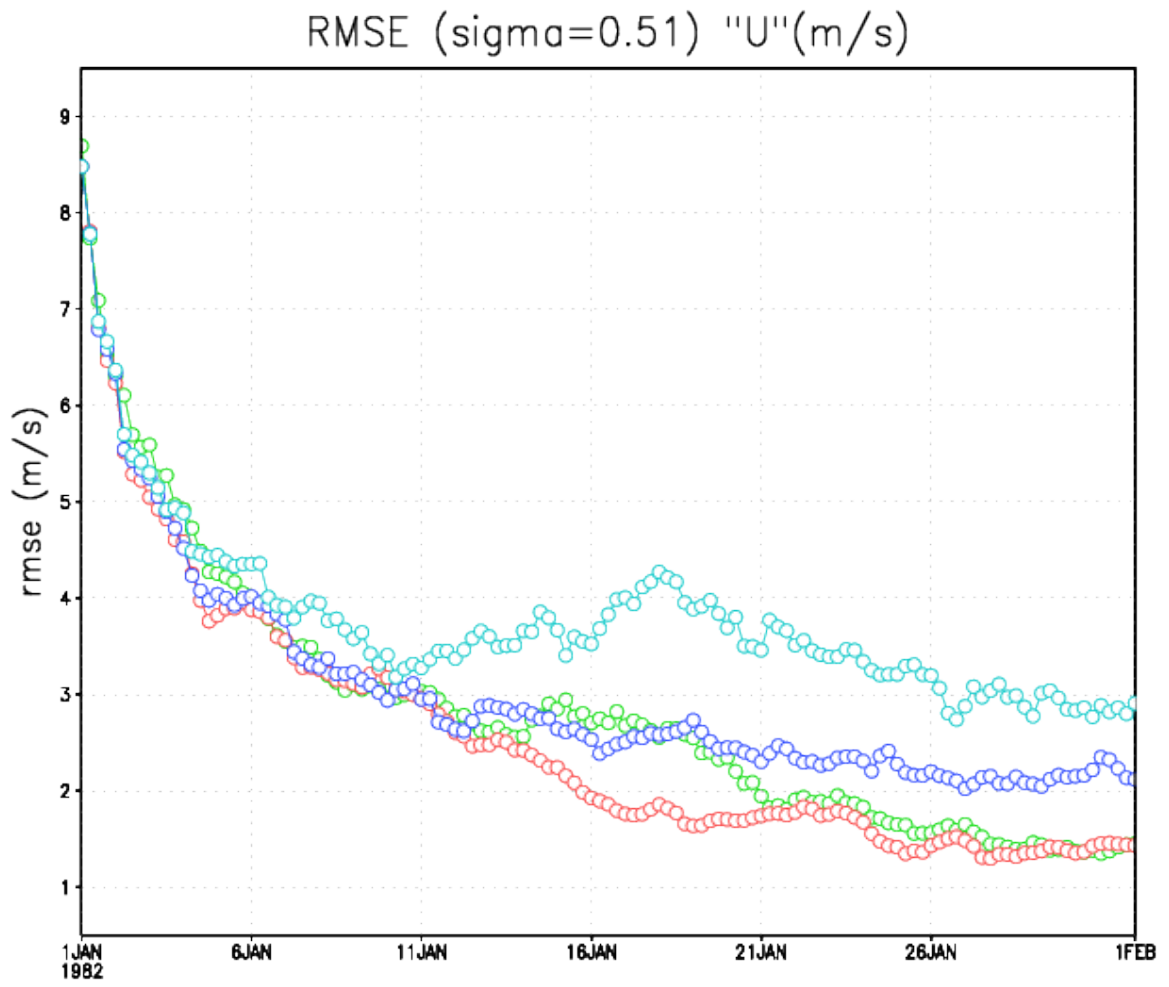


図 7: 鉛直第 4 層における東西風の解析二乗平均平方根誤差 (RMSE)。それ以外は図 6 に同じ。

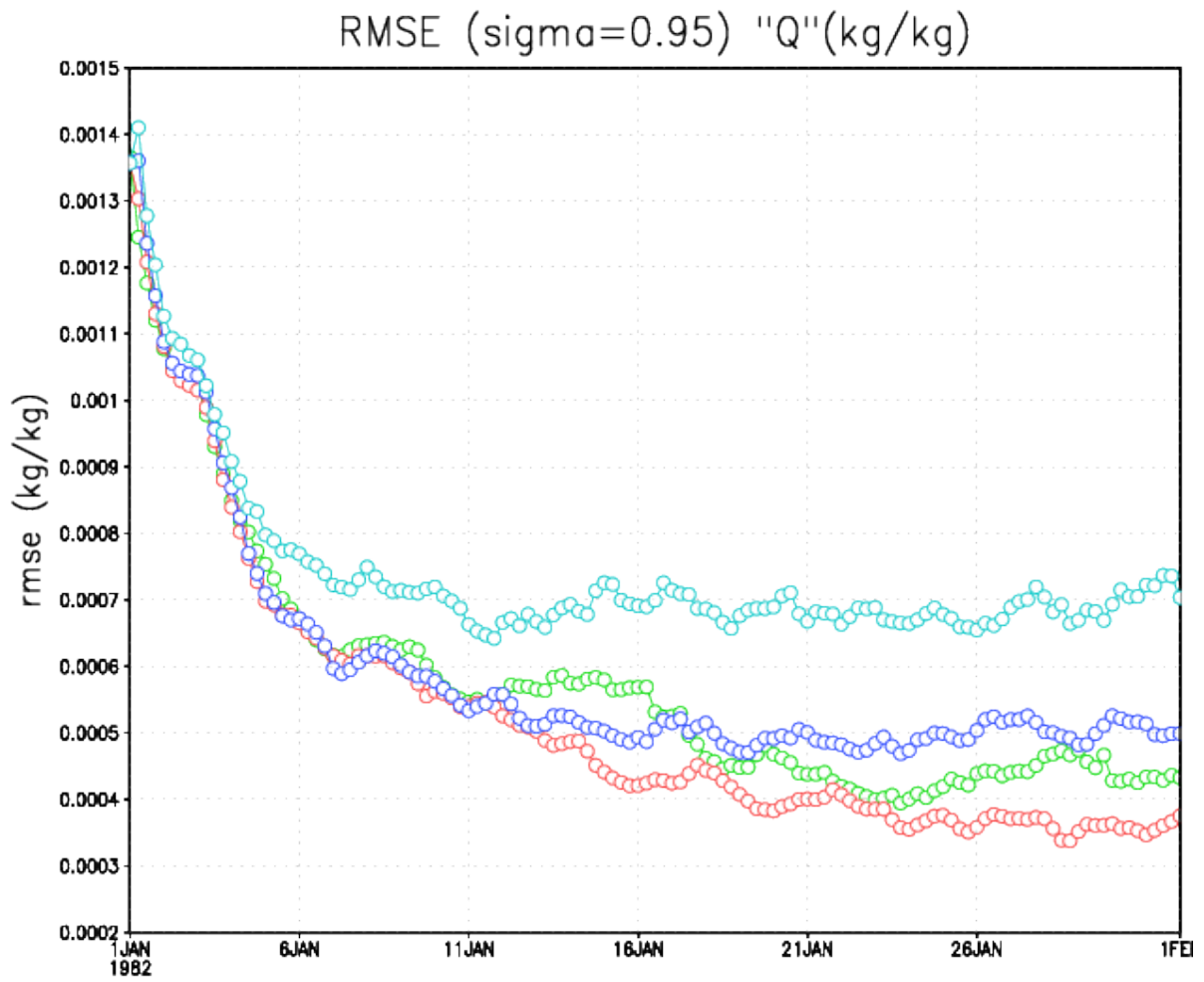


図 8: 鉛直第 2 層における比湿の解析二乗平均平方根誤差 (RMSE). それ以外は図 6 に同じ.

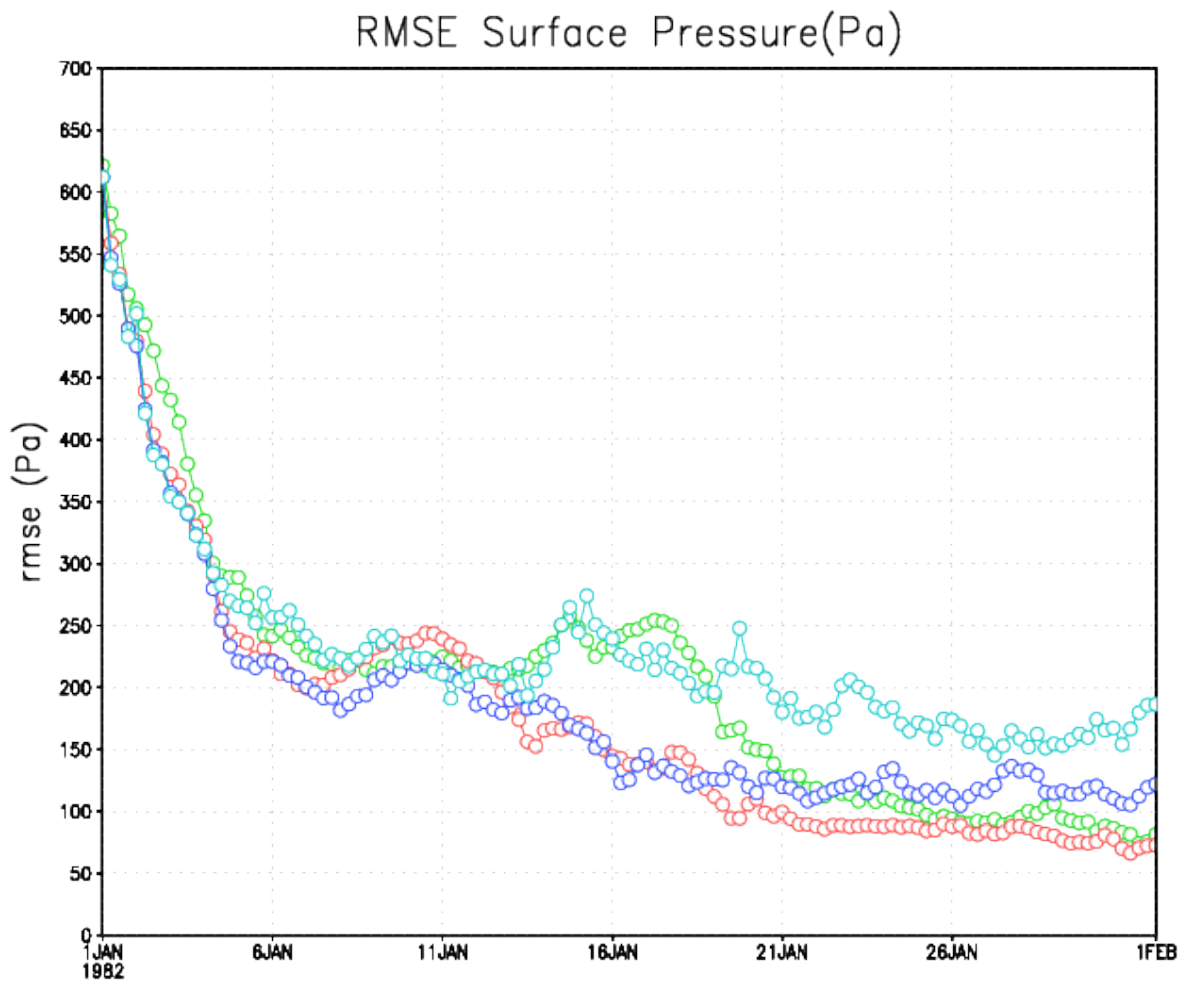


図 9: 地表面気圧の解析二乗平均平方根誤差 (RMSE). それ以外は図 6 に同じ.

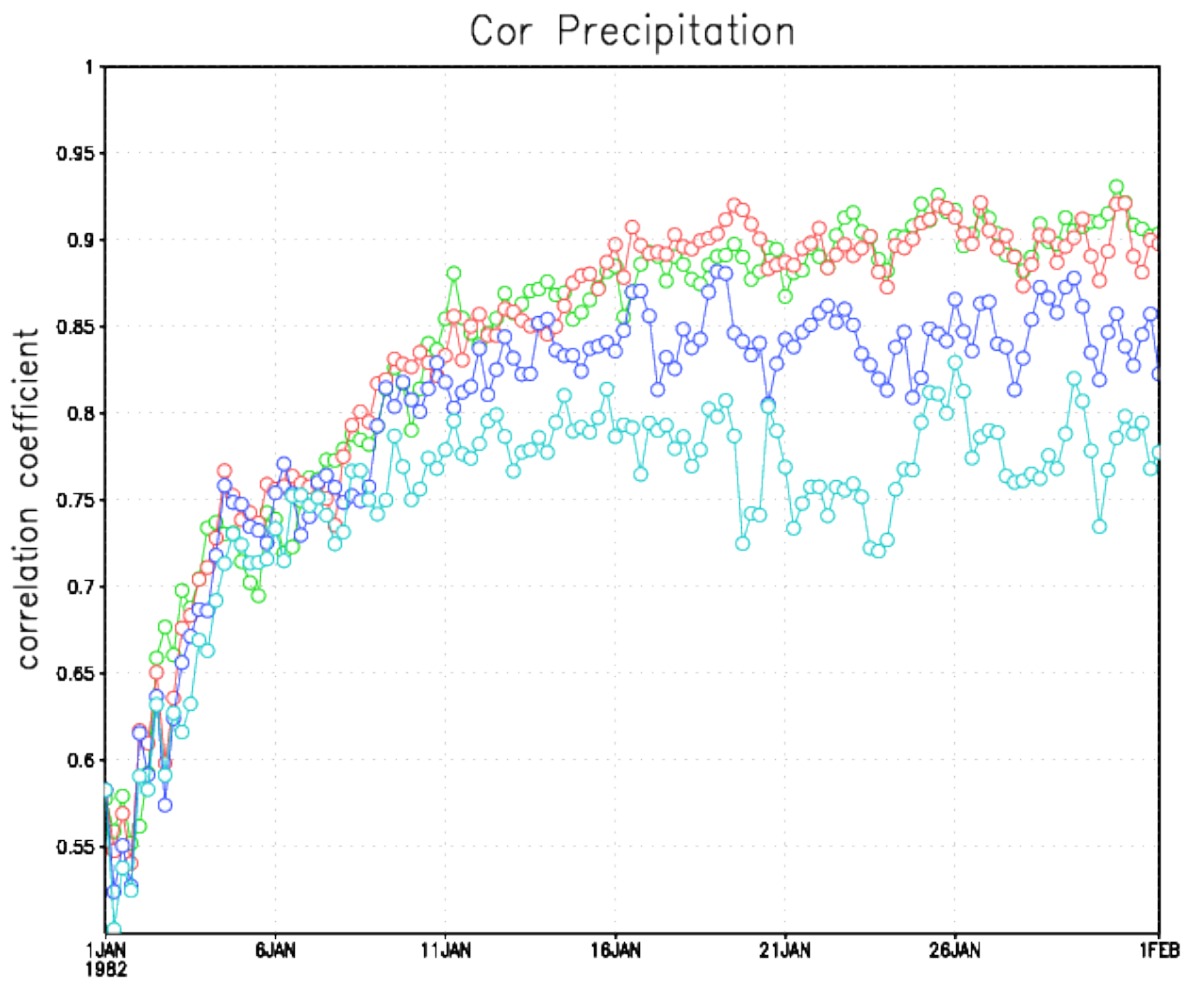


図 10: 全球での 6 時間積算降水量の真値に対する相関係数. 凡例等は図 6 に同じ.

RMSE (sigma=0.51) "T"(K) LETKF

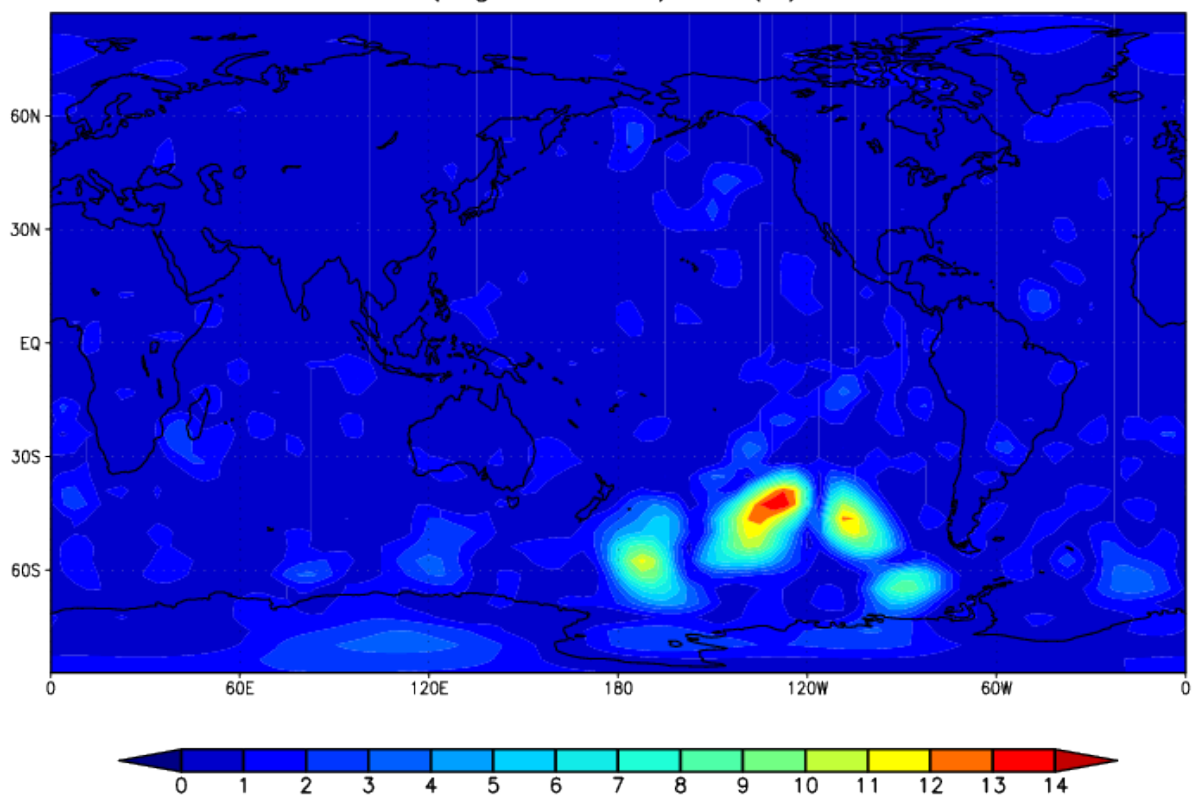


図 11: 1982 年 1 月 15 日 12 時における鉛直第 4 層における LETKF の解析の気温の RMSE. 陰影が RMSE の値を表しており, 暖色系ほど値が大きいことを示す.

RMSE (sigma=0.51) "T"(K) EnKS

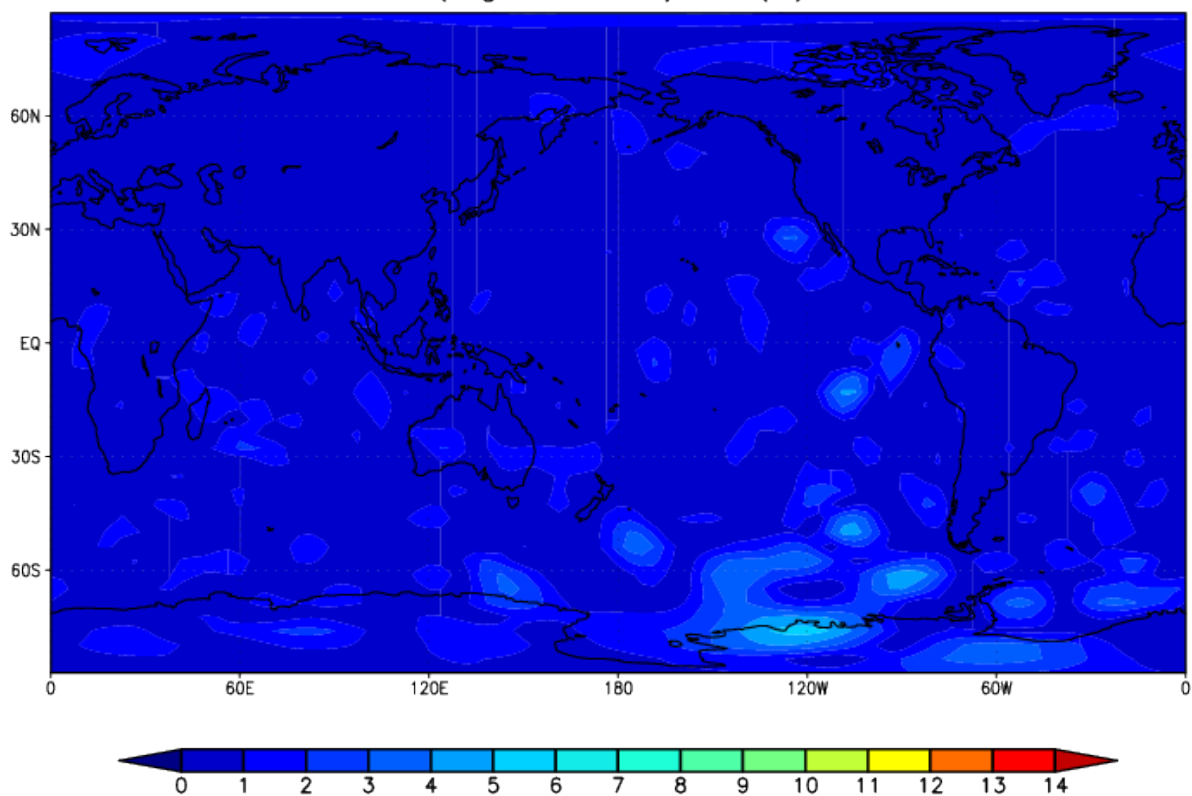


図 12: 1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 1 回) の解析の気温の RMSE. 陰影の設定は図 11 と同じ.

RMSE (sigma=0.51) "T"(K) EnKS(iter=2)

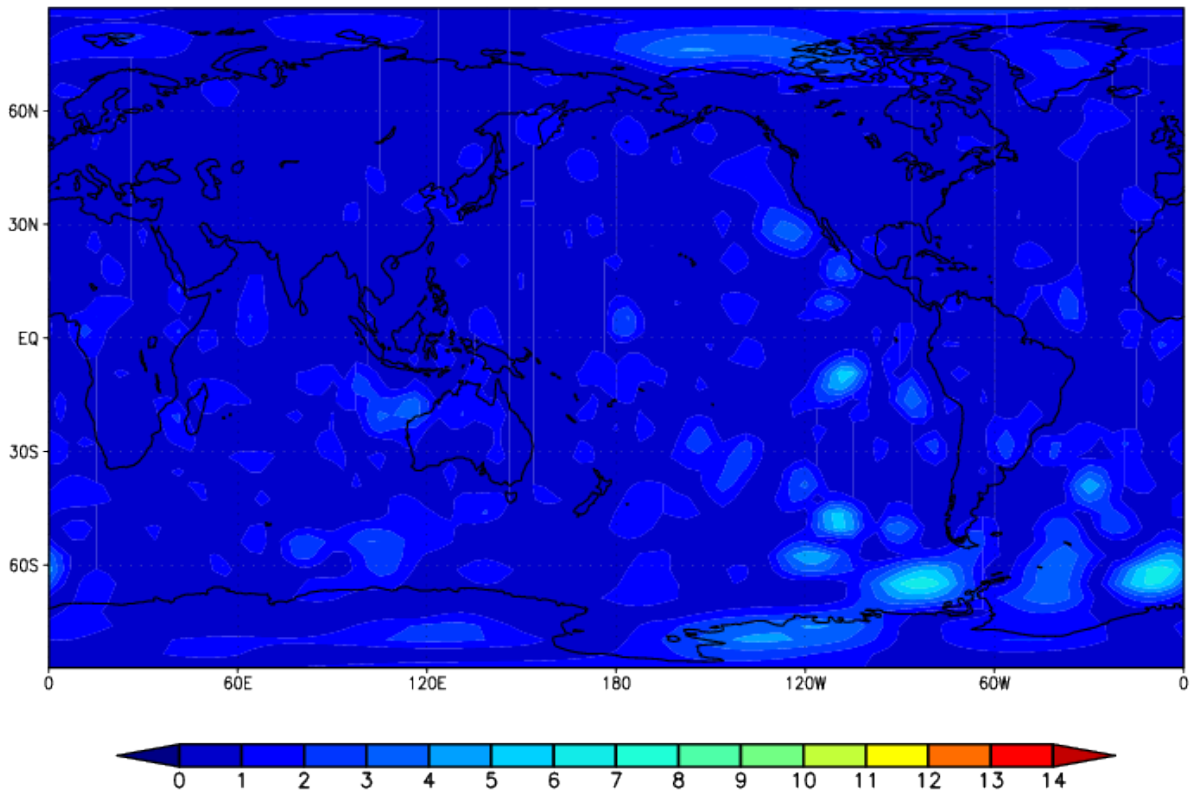


図 13: 1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 2 回) の解析の気温の RMSE. 陰影の設定は図 11 と同じ.

RMSE (sigma=0.51) "T"(K) EnKS(iter=3)

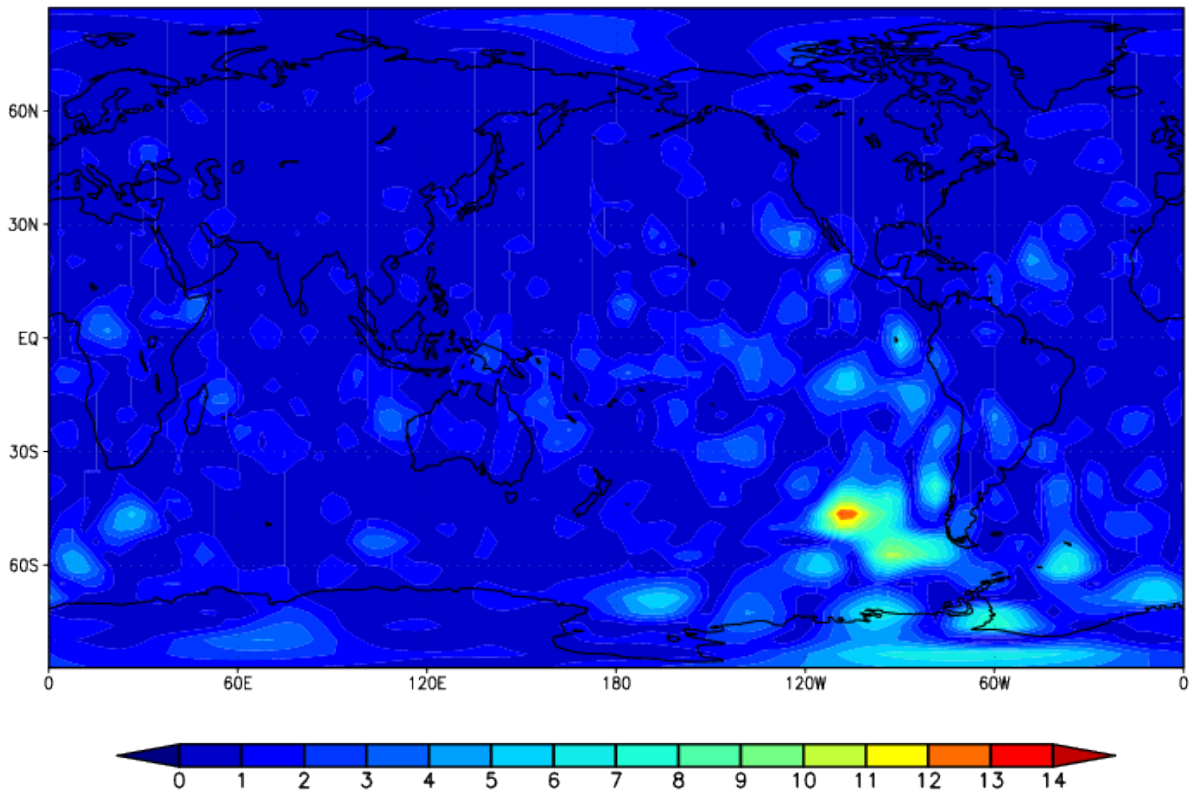


図 14: 1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 3 回) の解析の気温の RMSE. 陰影の設定は図 11 と同じ.

SPRD (sigma=0.51) "T"(K) LETKF

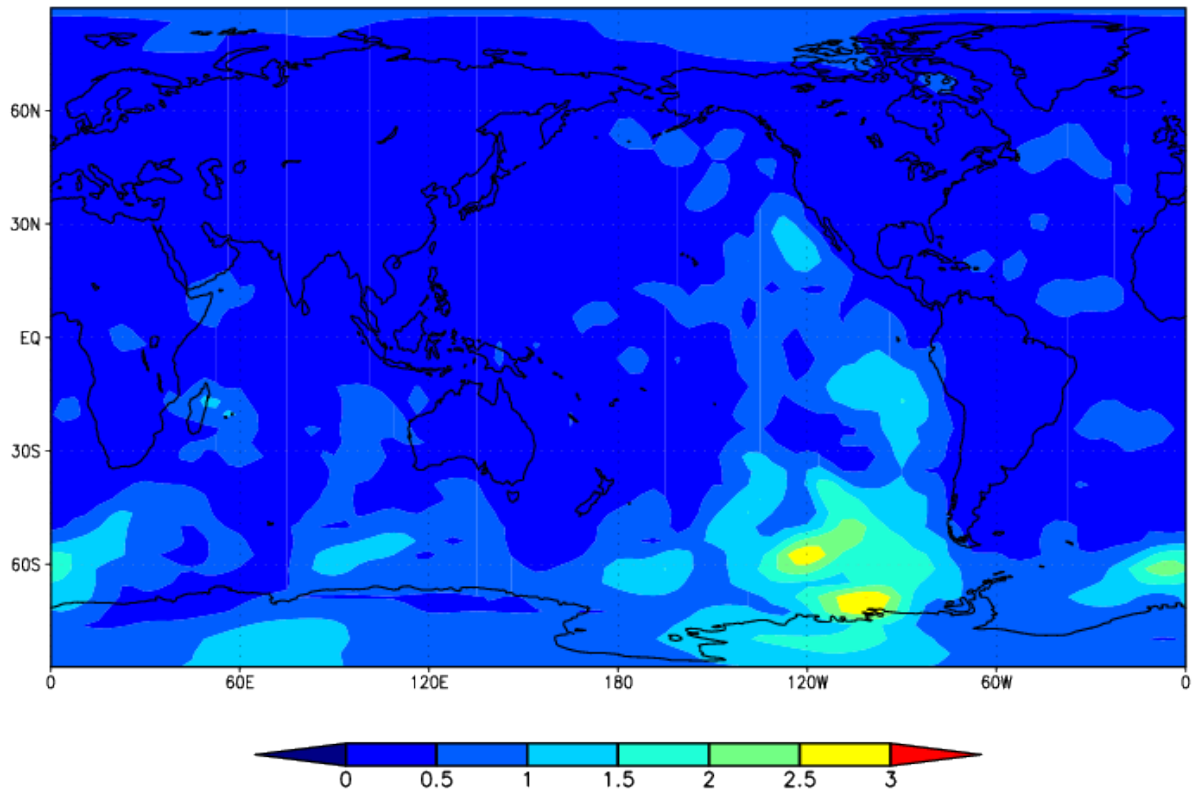


図 15: 1982 年 1 月 15 日 12 時における鉛直第 4 層における LETKF の解析の気温の Spread. 陰影が spread の値を表しており, 暖色系ほど値が大きいことを示す.

SPRD (sigma=0.51) "T"(K) EnKS

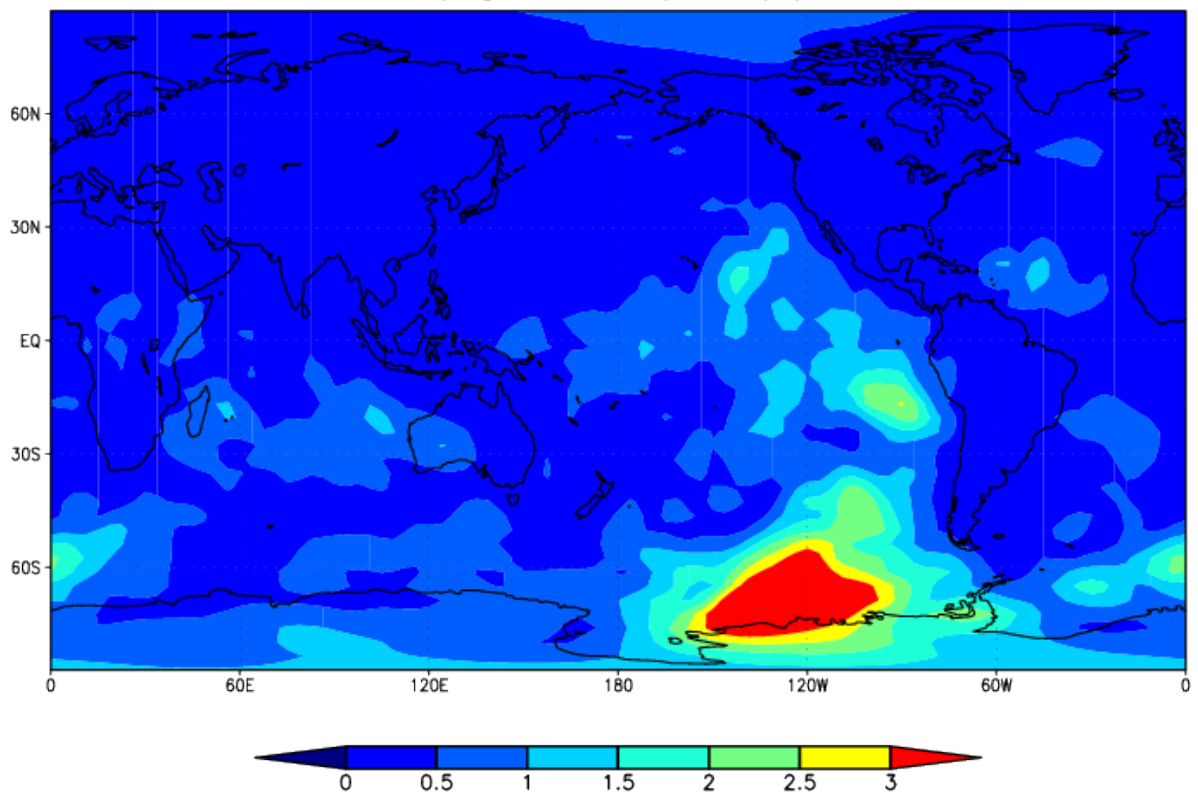


図 16: 1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 1 回) の解析の気温の Spread. 陰影の設定は図 15 に同じ.

SPRD (sigma=0.51) "T"(K) EnKS(iter=2)

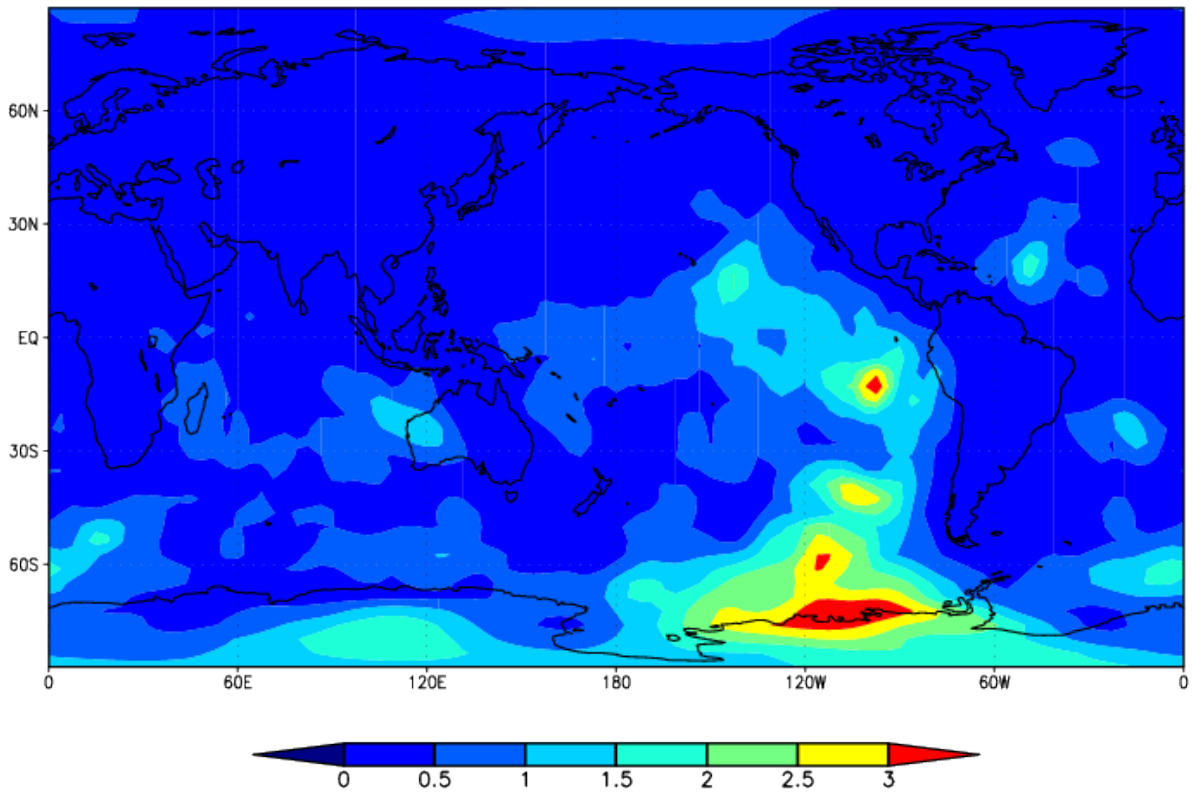


図 17: 1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 2 回) の解析の気温の Spread. 陰影の設定は図 15 に同じ.

SPRD (sigma=0.51) "T"(K) EnKS(iter=3)

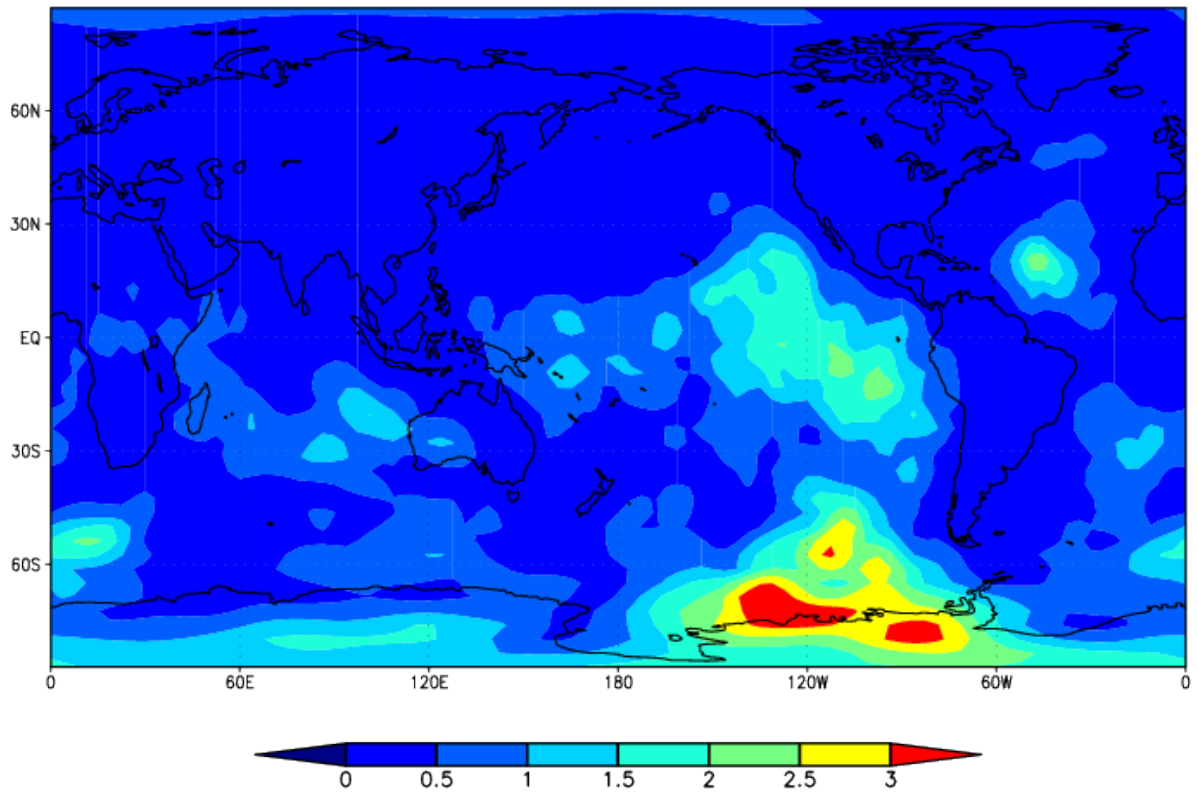


図 18: 1982 年 1 月 15 日 12 時における鉛直第 4 層における No-cost EnKS (繰り返し 3 回) の解析の気温の Spread. 陰影の設定は図 15 に同じ.

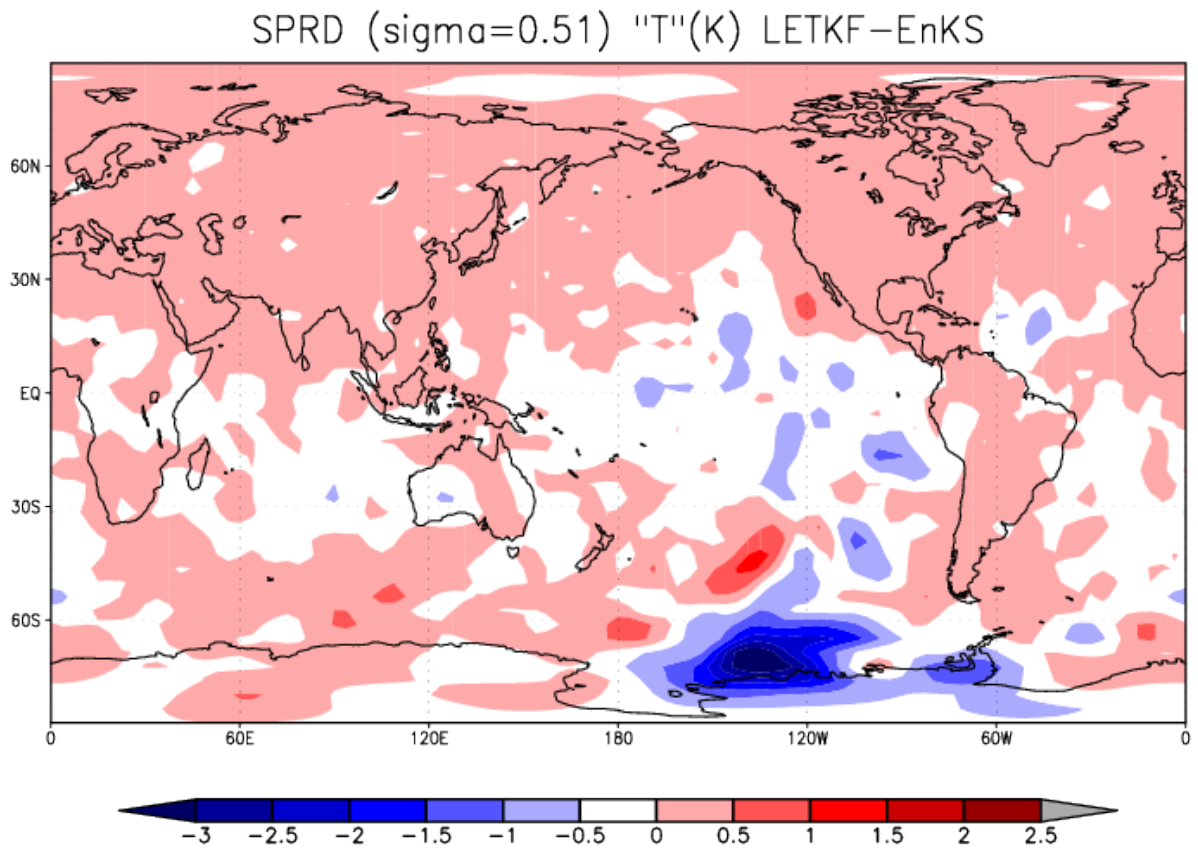


図 19: 1982 年 1 月 15 日 12 時における鉛直第 4 層における LETKF の解析 spread と No-cost EnKS (繰り返し 1 回) の解析 spread の差分。陰影は暖色系になるほど LETKF のほうが spread が大きいことを示し，寒色系になるほど No-cost EnKS のほうが spread が大きいことを示す。

SPRD (sigma=0.51) "T"(K) LETKF-EnKS(iter=2)

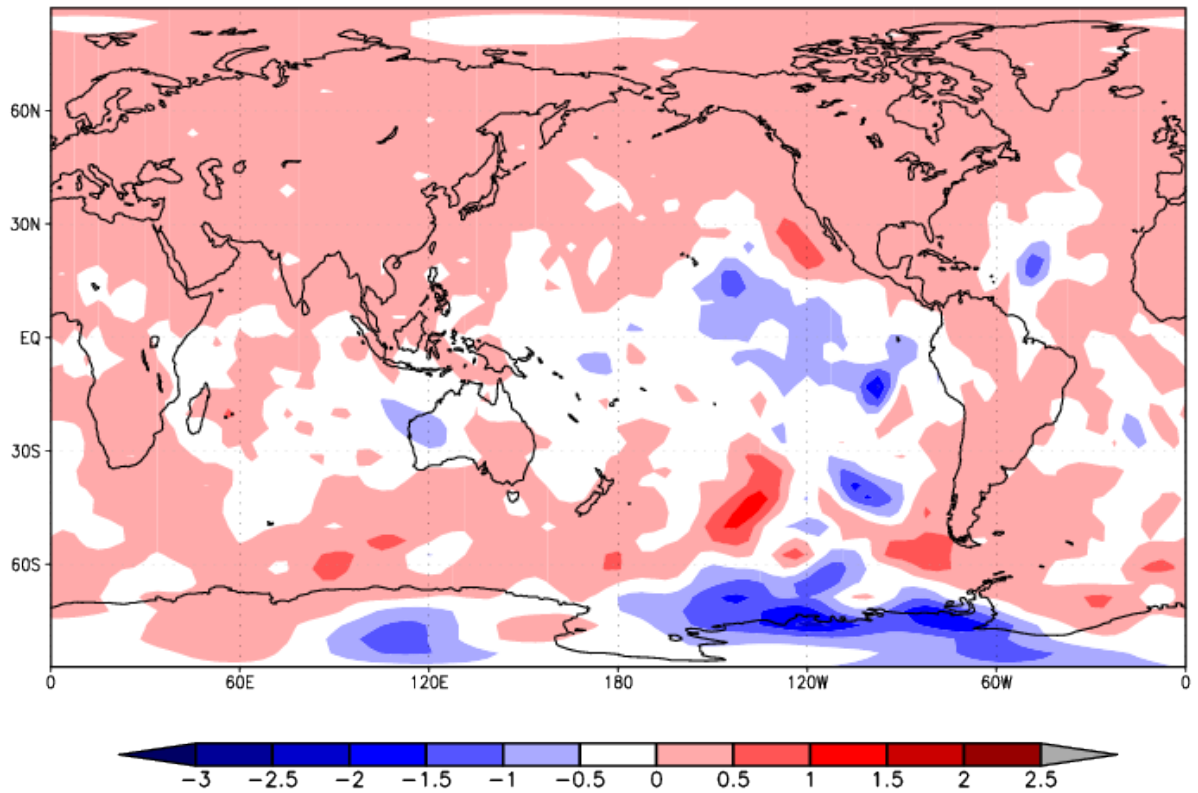


図 20: 1982 年 1 月 15 日 12 時における鉛直第 4 層における LETKF の解析 spread と No-cost EnKS (繰返し 2 回) の解析 spread の差分. 陰影の設定は図 19 に同じ.

SPRD (sigma=0.51) "T"(K) LETKF-EnKS(iter=3)

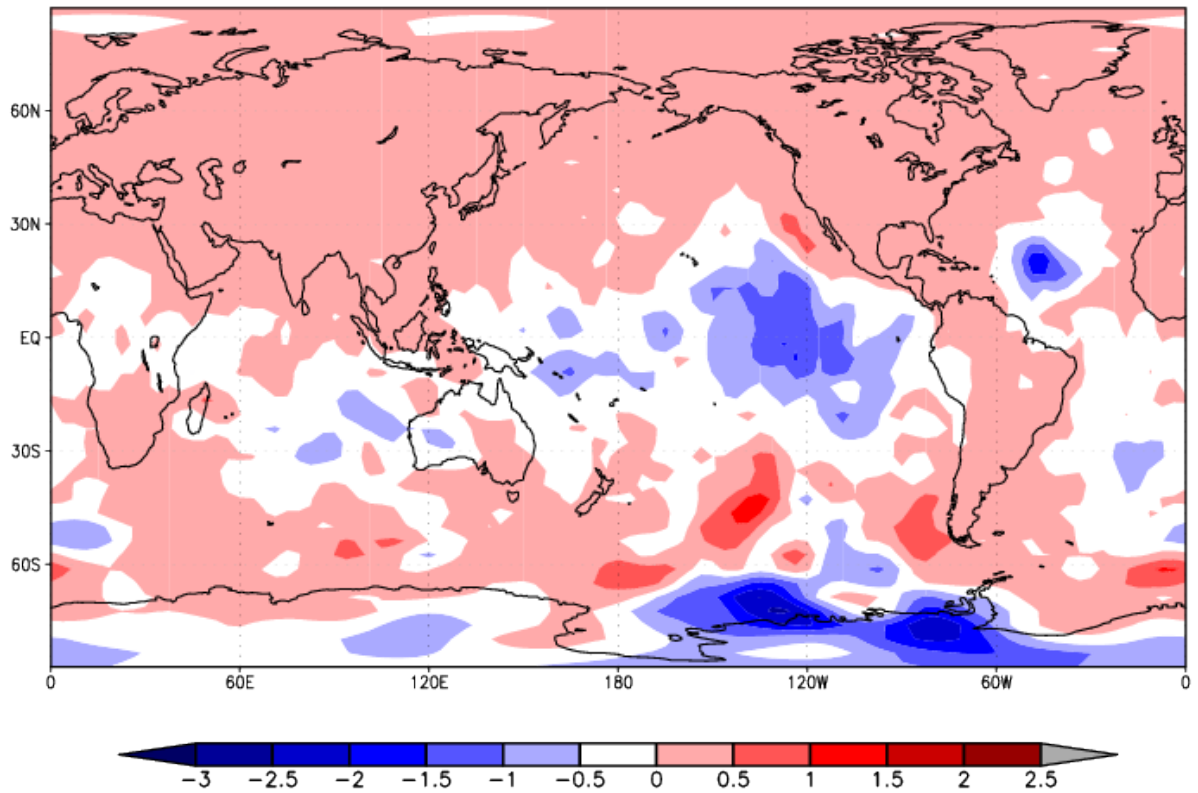


図 21: 1982 年 1 月 15 日 12 時における鉛直第 4 層における LETKF の解析 spread と No-cost EnKS (繰り返し 3 回) の解析 spread の差分. 陰影の設定は図 19 に同じ.

SPEEDY BGERR STDEV NMC_METHOD (U,sigma=0.51)

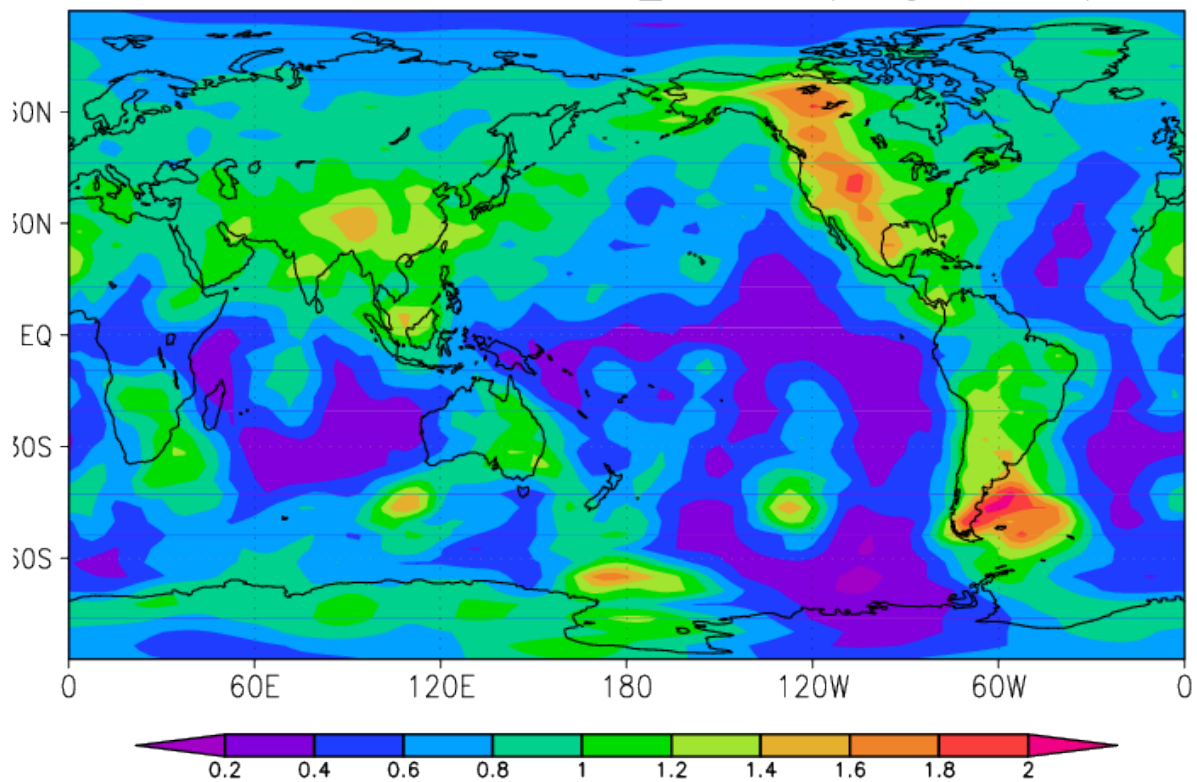


図 22: NMC 法で作成した, 第4層における背景誤差の東西風の標準偏差.

SPEEDY BGERR STDEV NMC_METHOD (T,sigma=0.51)

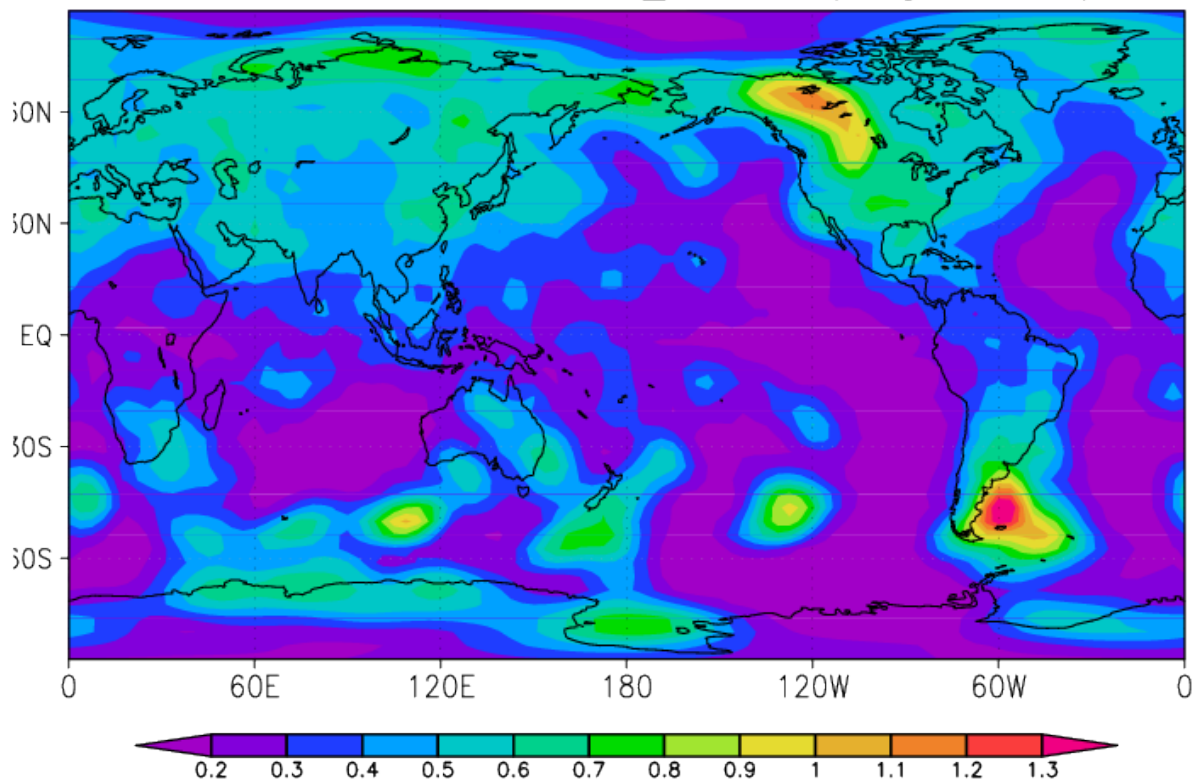


図 23: NMC 法で作成した, 第 4 層における背景誤差の気温の標準偏差.

SPEEDY BGERR STDEV ENSEMBLE (U,sigma=0.51)

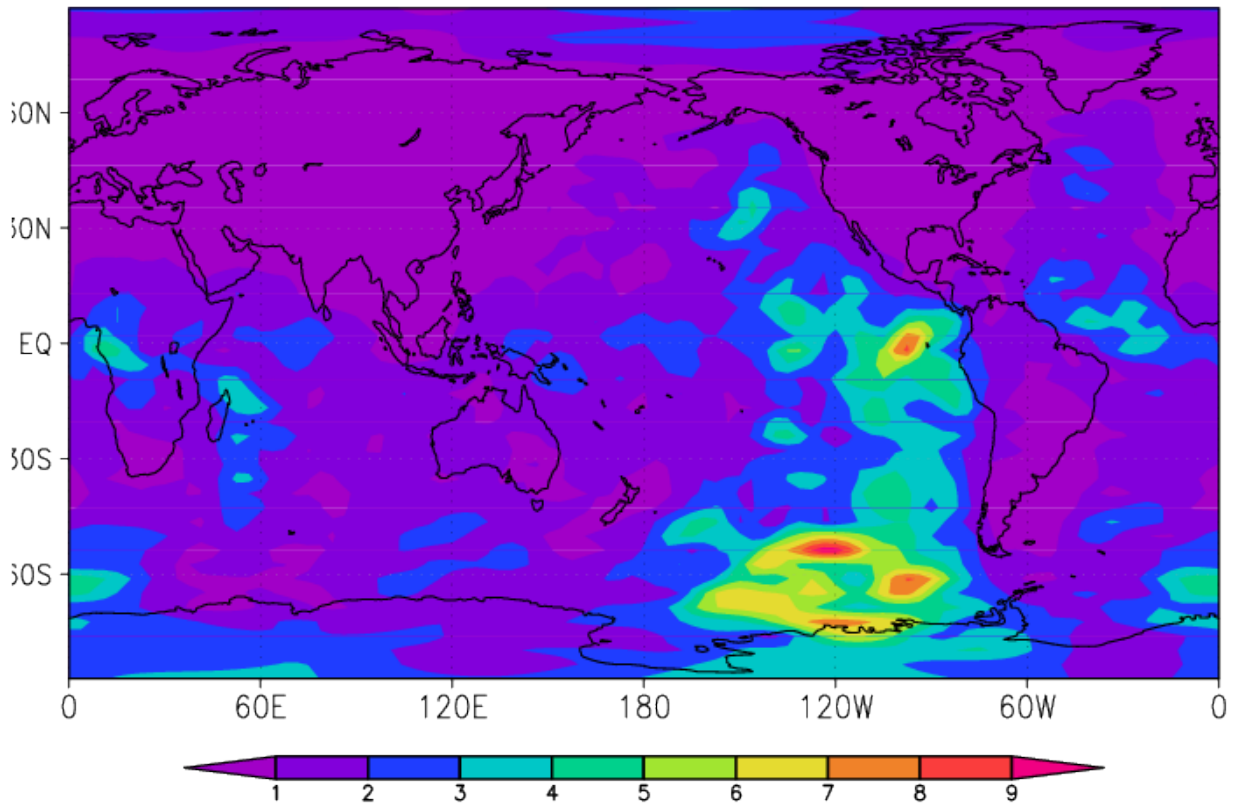


図 24: 1982 年 1 月 15 日 12 時における解析アンサンブルで作成した, 第 4 層における背景誤差の東西風の標準偏差.

SPEEDY BGERR STDEV ENSEMBLE (T,sigma=0.51)

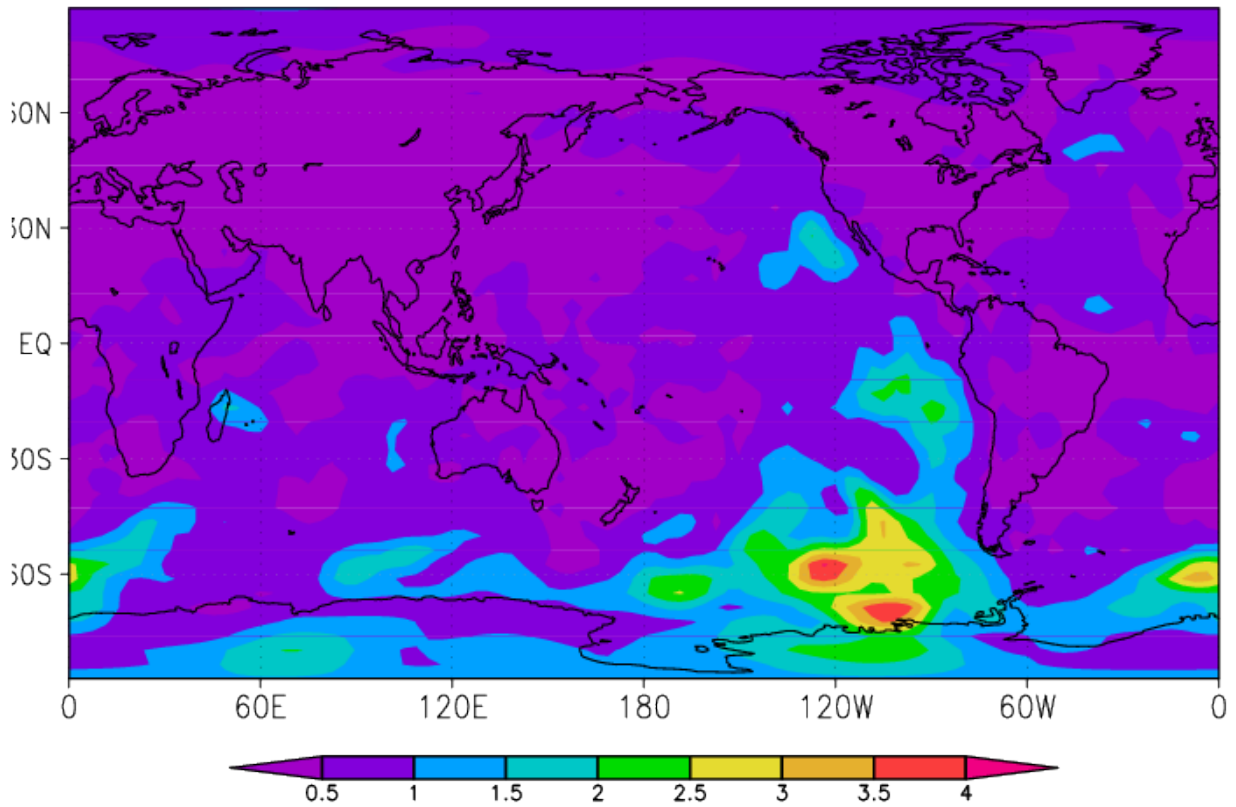


図 25: 1982 年 1 月 15 日 12 時における解析アンサンブルで作成した, 第 4 層における背景誤差の気温の標準偏差.

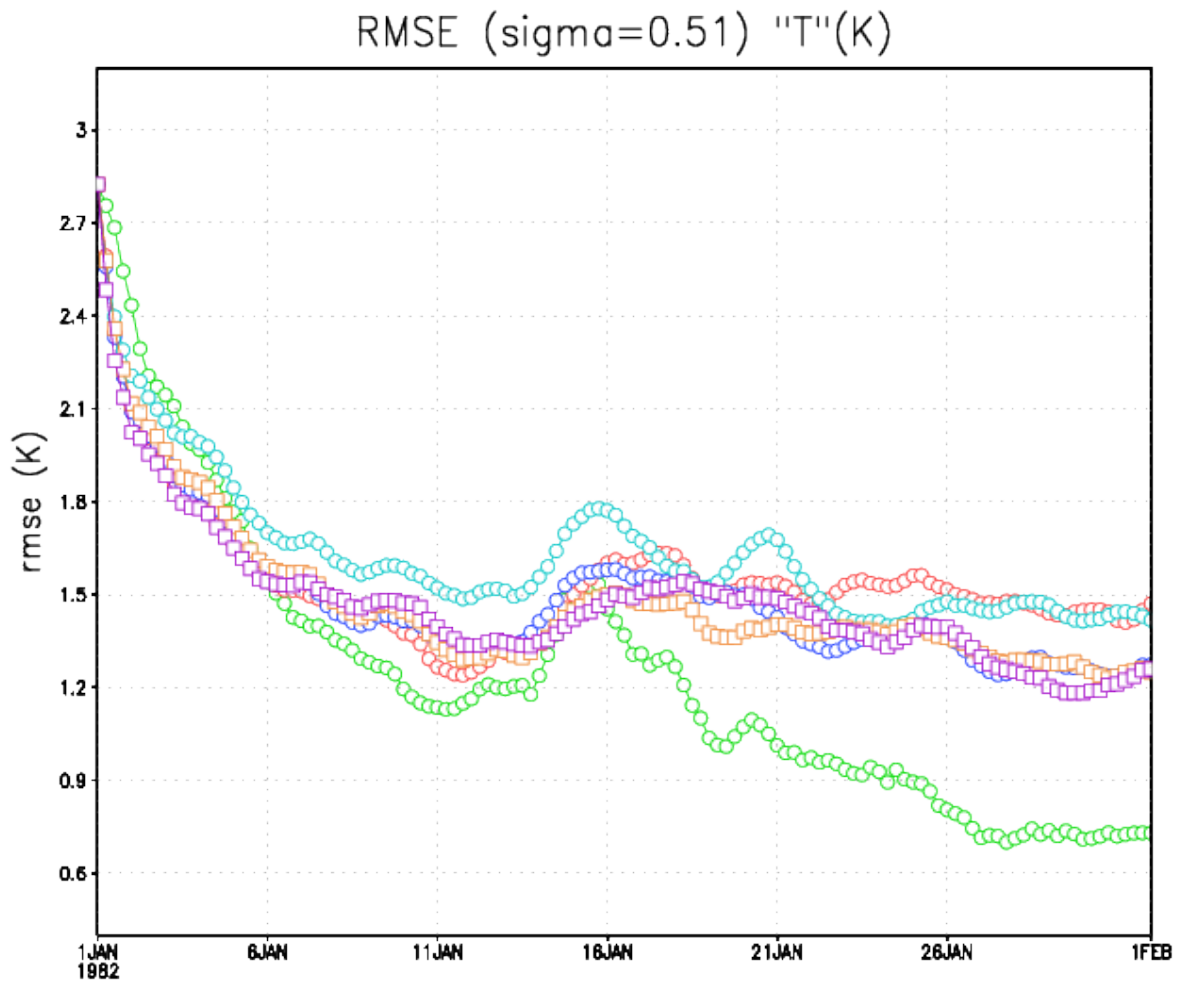


図 26: 鉛直第 4 層における気温の解析 RMSE. 縦軸が RMSE の値で, 横軸が時系列を表している. それぞれ, 緑色の円が通常の LETKF, 赤色の円が通常の LETKF による解析アンサンブルによる背景誤差のみを用いた 3DVAR, 青色の円が Hybrid-3DVAR, 水色の円が通常の NMC 法による背景誤差のみを用いた 3DVAR, 橙色の四角形が No-cost EnKS による解析アンサンブルを用いた Hybrid-3DVAR, 紫色の四角形が LETKF のメンバー数を 80 に増やしたときの Hybrid-3DVAR を示す.

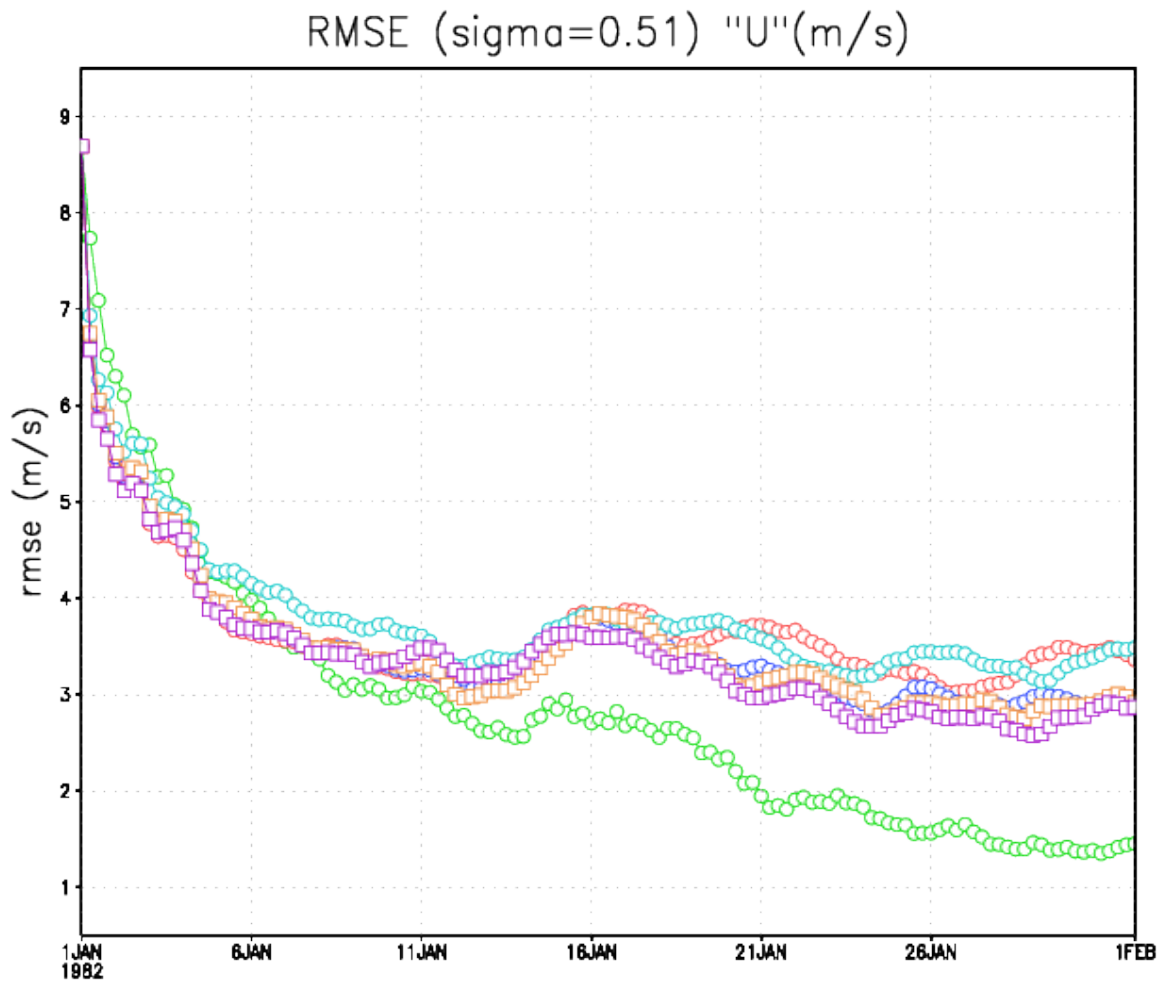


図 27: 鉛直第 4 層における東西風の解析 RMSE. 縦軸が RMSE の値で, 横軸が時系列を表している. 凡例は図 26 に同じ.